UNIVERSITY OF EDINBURGH

College of Science and Engineering

School of Chemistry

Peptide nucleic acid-encoded libraries for microarray-based enzymatic high-throughput screening

By

Delphine Pouchain

Doctor of Philosophy

June 2009

UNIVERSITY OF EDINBURGH

ABSTRACT

College of Science and Engineering

School of Chemistry

Doctor of Philosophy

Peptide nucleic acid-encoded libraries for microarray-based enzymatic high-throughput screening

By Delphine Pouchain

Protein kinases represent one of the largest families of enzymes. Due to their crucial roles in many cellular processes, protein kinases have become one of the major drug targets for pharmaceutical companies. In order to be able to determine selective substrates, elucidate pathway functionality and design potential kinase inhibitors, it is important to determine kinase substrate specificity. A 10,000-member split and mix peptide nucleic acid (PNA)-encoded peptide library was used to establish the substrate specificity of three different protein tyrosine kinases, using a DNA microarray and a Cy3-fluorescently labelled secondary anti-phosphotyrosine antibody. This approach was proven to be a powerful tool as several peptides ("hits") were rapidly identified as good substrates for the three kinases.

Split and mix combinatorial libraries can generate large peptide libraries consisting of potentially millions of compounds, but the main disadvantage is the deconvolution process required to identify active individual peptides from mixtures. The aim of the second part of the project was to build a PNA-encoded positional scanning library of FRET-based tetrapeptides to carry out microarray-based proteolytic profiling. As proof-of-principle, a 625-member library was first synthesised on solid phase using a split and mix methodology, and successfully used to profile three different proteases. The concept was then extended to the synthesis of a larger positional scanning library containing 50,625 different tetrapeptides encoded for by just 60 PNA tags.

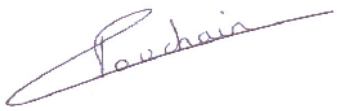
DECLARATION

I, Delphine Pouchain, declare that the thesis entitled Peptide nucleic acid-encoded libraries for microarray-based enzymatic high-throughput screening and the work presented in it are my own.

I confirm that:

• this work was done wholly or mainly while in candidature for a research degree at this University;
• where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
• where I have consulted the published work of others, this is always clearly attributed;
• where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
• I have acknowledged all main sources of help;
• where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Date: 17/09/09

Signed: 

PREFACE

The research work described in this thesis was carried out under the supervision of Prof. Mark Bradley at the University of Southampton (Oct. 2004 – Feb. 2005) and the University of Edinburgh (March 2005 – Sept. 2008).

Part of the work presented herein has been published as:

- "A 10,000 member PNA-encoded peptide library for profiling tyrosine kinases", Pouchain, D.; Díaz-Mochón, J. J.; Bialy, L.; Bradley, M. *ACS Chemical Biology*, **2007**, *2*, 810-818.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Professor Mark Bradley, for all his help and support throughout my PhD, and for giving me this great opportunity of carrying out a PhD in his research group.

I would like to thank all members of the Bradley group (both past and present) who made this time so special and enjoyable. They are too numerous to list individually, but in particular I would like to thank all members of the Combinatorial Centre of Excellence (Christophe, Eric, Luciano, Alessandra, Gianluca) and show my recognition to Juan Jo who provided me with valuable advice and help throughout my PhD. Special thanks to Juan Jo, Laurent, Edgar, Eric, Nina and Song for sharing experiences on the PNA synthesis.

Many thanks go to Christophe, Maria L., Maria M., Edgar, Song, Adam, Salvo, Rosario and Géraldine for their continuous smiles, and happy times outside the lab. A special “thank you” goes to Géraldine for her friendship and her help while I was writing up my thesis in France.

Finally, I would like to thank all my friends from France (Amiens, Montpellier), especially Céline and Pearl who are fantastic friends, and my family for their continual support all along my studies, especially my parents, Corentin and Amandine.

TABLE OF CONTENTS

ABBREVIATIONS

Amino acids

Three-letter code (one-letter code)

Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
Cys (C)	cysteine
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
His (H)	histidine
Ile (I)	isoleucine
Leu (L)	leucine
Lys (K)	lysine
Met (M)	methionine
Phe (F)	phenylalanine
Pro (P)	proline
Ser (S)	serine
Thr (T)	threonine
Trp (W)	tryptophan
Tyr (Y)	tyrosine
Val (V)	valine

General

AA	amino acid
Abl	abelson tyrosine kinase
Abu	amino butyric acid
Ac	acetyl
ACC	7-amino-4-carbamoylmethylcoumarin
Aloc	allyloxycarbonyl
AP	alkaline phosphatase
aPK	atypical protein kinase
AMC	7-amino-4-methylcoumarin
Ar	aryl (NMR)
ATP	adenosine triphosphate
Azoc	azidomethyloxycarbonyl
Bcr	breakpoint cluster region

Bhoc	benzhydryloxycarbonyl
BLAST	basic local alignment search tool
Boc	*tert*-butoxycarbonyl
BOP	benzotriazolyl-*N*-oxytrisdimethylaminophosphonium hexafluorophosphate
br	broad (NMR)
BSA	bovine serum albumin
Bzl	benzoyl
Cbz	carbobenzyloxy
cDNA	complementary deoxyribonucleic acid
CML	chronic myelogenous leukemia
d	doublet (NMR)
Dabsyl	4-(4-dimethylaminophenylazo)benzoic acid
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
dd	double doublet (NMR)
Dde	*N*-[1-(4,4-dimethyl-2,6-dioxo*cyclo*hex-1-ylidene)ethyl]
DIC	diisopropylcarbodiimide
DIPEA	*N, N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	dimethylsulfoxide
DMT	dimethoxytrityl
DNA	deoxyribonucleic acid
dok1	docking protein 1
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDT	1,2-ethanedithiol
EDTA	ethylenediaminetetraacetatic acid
EGFR	epidermal growth factor receptor
ELSD	evaporative light scattering detector
ePK	eukaryotic protein kinase
ESI	electrospray ionisation
ESAC	encoded self-assembling chemical
FISH	fluorescence *in situ* hybridisation
Fmoc	9-fluorenylmethoxycarbonyl
FRET	fluorescence resonance energy transfer
GPCR	G protein coupled receptor
HATU	*O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	*O*-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
Her2	human epidermal growth factor 2
HF	hydrofluoric acid
HIV	human immunodeficiency virus
HOAt	1-hydroxy-7-azabenzotriazole

HOBt	1-hydroxy-1*H*-benzotriazole
HPLC	high performance liquid chromatography
HTS	high-throughput screening
IgG	immunoglobulin G
IR	infrared
J	scalar coupling constant (NMR)
JNK	c-Jun amino-terminal kinase
m	multiplet (NMR) or medium (IR)
MALDI-TOF	matrix-assisted laser desorption/ionisation time-of-flight
Mdm2	mouse double minute 2
Me	methyl
Mmt	monomethoxytrityl
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MYOD1	myoblast determination protein 1
NEM	*N*-ethylmorpholine
NHS	*N*-hydroxysuccinimidyl
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
Pbf	2,2,4,6,7-pentamethyldihydro-benzofuran-5-sulfonyl
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDMS	poly(dimethylsiloxane)
PEG	polyethyleneglycol
PEGA	polyethylene glycol dimethyl acrylamide
PKA	protein kinase A
PLL	photolabile linker
PNA	peptide nucleic acid
PTN6	tyrosine-protein phosphatase non-receptor type 6
Pur	purine
PyBOP	benzotriazol-1-yloxytri(pyrrolidino)-phosphonium hexafluorophosphate
q	quartet (NMR)
quint.	quintuplet (NMR)
RFU	relative fluorescence units
RNA	ribonucleic acid
s	singlet (NMR) or strong (IR)
SAM	self-assembled monolayer
SDS	sodium dodecyl sulfate
SNP	single-nucleotide polymorphism
SPPS	solid phase peptide synthesis
SSC	sodium chloride/sodium citrate

StdDev	standard deviation
t	triplet (NMR)
TAMRA	tetramethyl-6-carboxyrhodamine
td	triplet of doublets (NMR)
TFA	trifluoroacetic acid
TFMSA	trifluoromethanesulfonic acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin layer chromatography
TPBS	Tween-20 phosphate buffered saline
T_m	melting temperature
TR	time-resolved
t_R	retention time
UV	ultraviolet
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
w	weak (IR)
Z	benzyloxycarbonyl
δ	chemical shift (NMR)

Chapter 1: Introduction

1.1 Peptide nucleic acids

1.1.1 Introduction

A number of issues related to conventional deoxyribonucleic acids (DNAs) or ribonucleic acids (RNAs) such as enzymatic degradation and mismatch discrimination have led to the development of numerous DNA (or RNA) analogues. Peptide nucleic acids (PNAs) have appeared as one very attractive synthetic DNA mimic, with the first synthesis reported by Nielsen in 1991.[1] Instead of the phosphate deoxyribose backbone of DNAs **1.1**, PNAs **1.2** possess an uncharged, achiral and acyclic polyamide backbone composed of *N*-(2-aminoethyl)glycine units. Purine (adenine **1.3** and guanine **1.4**) and pyrimidine (cytosine **1.5** and thymine **1.6**) bases are attached to the backbone through a methylene carbonyl spacer (Figure 1.1).[2]

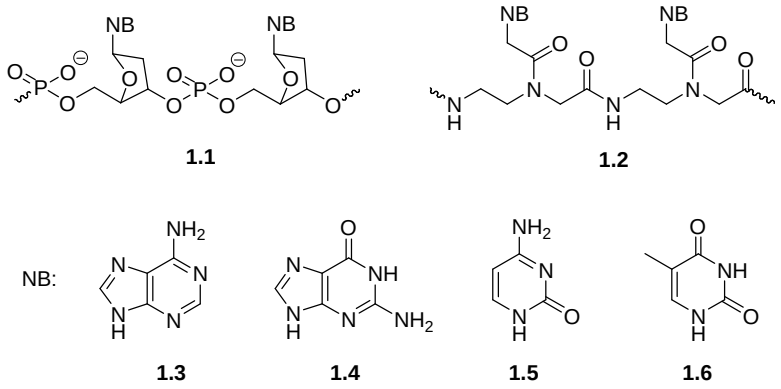


Figure 1.1: Chemical structures of a DNA and a PNA, where NB is one of the four natural nucleobases

1.1.2 Properties

PNAs can bind specifically to complementary single-stranded DNA, RNA or PNA to form very stable duplexes based on Watson-Crick base pairing.[3] By convention, like peptides, PNA sequences are written from the *N*- to the *C*-terminus. Duplexes can be formed both in the anti-parallel and parallel orientations but the anti-parallel orientation (*N*-terminus of PNA facing the 3'-end of DNA) is strongly preferred

(Figure 1.2).[2] Due to the neutral amide backbone of PNAs, hybridisation can occur under low salt conditions, and the absence of electrostatic repulsions confers a higher thermal stability of PNA/DNA duplexes (Figure 1.2) than DNA/DNA duplexes. Thus, the melting temperature (T_m, temperature at which 50% of a duplex is dissociated into single strands) of a 15-mer PNA/DNA duplex is about 70 °C whereas a corresponding 15-mer DNA/DNA duplex has a T_m of about 55 °C. A much better mismatch discrimination is observed, with a single-base mismatch in a 15-mer PNA/DNA duplex reducing its T_m by 15 °C on average, in comparison to a decrease of 11 °C for the T_m of a corresponding DNA/DNA duplex.[3]

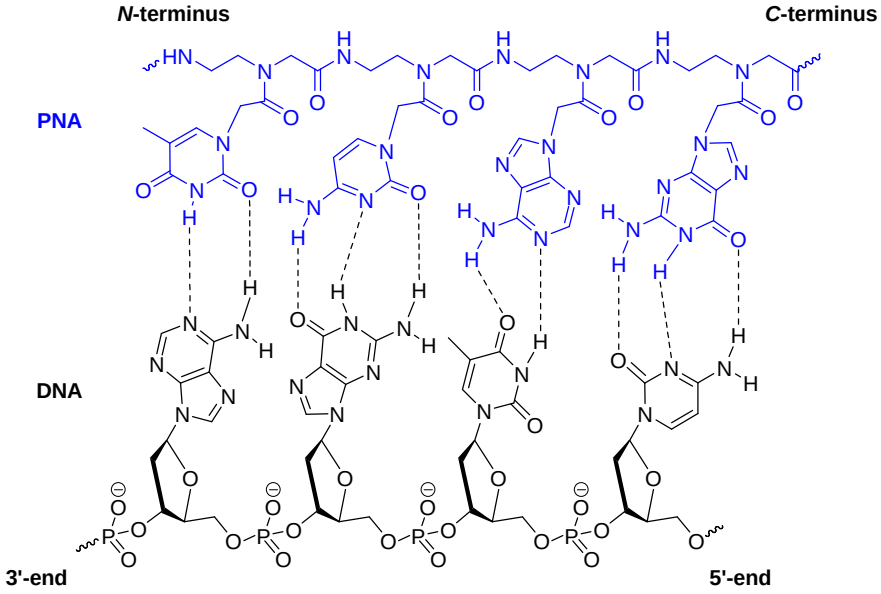


Figure 1.2: General structure of a PNA/DNA duplex, in the anti-parallel mode[4]

PNAs can bind to a double-stranded DNA to form stable triplexes, and different types of structures have been reported.[5] For example, PNA-$(DNA)_2$ triplexes, where a PNA binds to a double-stranded DNA through Hoogsteen base-pairing, are observed in the case of cytosine-rich PNAs and guanine/cytosine-rich DNA duplexes (Figure 1.3B).[6] However, under low salt concentrations and with a purine-rich DNA duplex, PNAs composed of only thymine and cytosine bases (so-called homopyrimidine PNAs) or containing a high pyrimidine/purine ratio prefer binding to a double-stranded DNA by strand displacement to form a highly stable triplex invasion complex (T_m > 85 °C for a 10-mer homopyrimidine PNA) (Figure 1.3C).[1, 7]

In that case, one PNA binds to the homopurine DNA strand following Watson-Crick base-pairing rules (by preference in the anti-parallel mode), while the other PNA binds to the PNA/DNA duplex through Hoogsteen hydrogen bonds (by preference in the parallel mode), thus forming an internal $(PNA)_2$-DNA triplex. The non-complementary single-stranded DNA is displaced and forms a so-called "P-loop structure" (Figure 1.3C).[7]

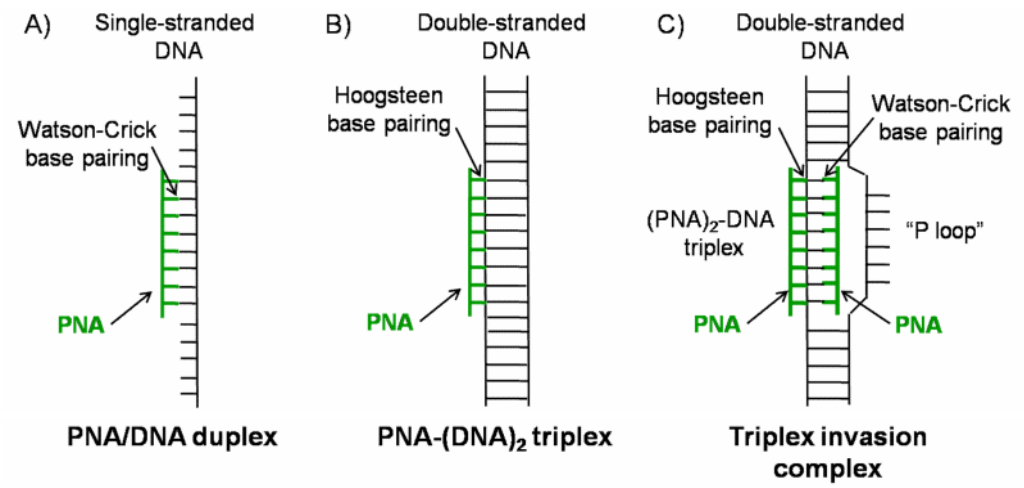


Figure 1.3: Schematic representations of (A) a PNA/DNA duplex, (B) a PNA-$(DNA)_2$ triplex and (C) a triplex invasion complex

However, the formation of triplex invasion complexes containing cytosine in their sequences is pH-dependent since protonated cytosines are required for Hoogsteen binding. This limitation can be avoided by connecting the two homopyrimidine PNA strands by a linker (*e.g.* an 8-amino-3,6-dioxaoctanoic acid linker[8] or a positively charged lysine-aminohexyl linker[9]) to form a so-called bis-PNA, and by replacing all the cytosine residues in the Hoogsteen strand with pseudoisocytosines (Figure 1.4).[8] Bis-PNAs are even more stable than triplex invasion complexes, and a T_m of 100 °C is observed for a 10-mer bis-PNA. The ability of PNAs to bind to a double-stranded DNA by strand displacement has been successfully used as a powerful tool in molecular biology and functional genomics.[10, 11] Nucleases and proteases do not recognise the chemical structure of PNAs. This resistance to enzymatic degradation extends the lifetime of PNAs both *in vivo* and *in vitro*,[12] and PNAs are stable over a

wide range of temperature and pH values. However PNAs cannot function as primers for DNA polymerases.[2]

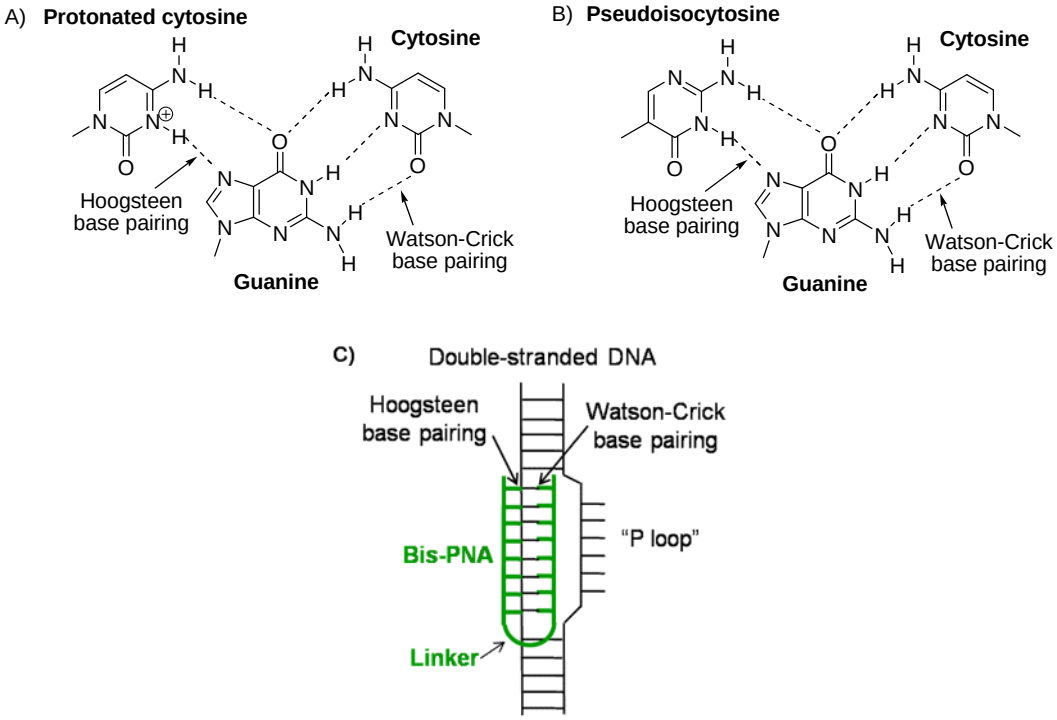


Figure 1.4: Triplex formation using (A) a protonated cytosine and (B) a pseudoisocytosine in the Hoogsteen strand; (C) schematic representation of a triplex invasion complex with a bis-PNA

1.1.3 Synthesis

Similarly to peptides, PNAs can be synthesised using standard solid phase synthesis procedures. PNAs are stable to strong acids and weak bases so that diverse protecting groups have been used for the synthesis of PNAs.[2] In two elemental strategies, the primary amino group of the backbone is protected with Fmoc[13] or Boc[14] groups, and the exocyclic nitrogens of the nucleobases with Bhoc or Cbz groups respectively. Fmoc/Bhoc- and Boc/Cbz-protected PNA monomers are commercially available, although very expensive. Other protections for the exocyclic amino groups of the nucleobases (*e.g.* Mmt, N_3, Aloc) have been reported, and the main protecting groups for the nucleobases are Cbz, Bhoc and Mmt groups.[15] Recently, the syntheses of Fmoc/bis-Boc-protected PNA monomers have also been published.[16, 17] However, the use of Boc/Cbz-protected PNA monomers implies several TFA deprotections during the PNA synthesis, and harsh acidic conditions (HF or TFMSA) are necessary

to remove the Cbz protecting groups (this is also the main issue in the case of Fmoc/Cbz-protected PNA monomers[18]). To protect the exocyclic amino groups of the nucleobases, the mild acid-labile Mmt group is preferred, and the synthesis of Fmoc/Mmt-protected PNA monomers has been reported.[19] Although Fmoc/Bhoc-, Fmoc/bis-Boc- and Fmoc/Mmt-protected PNA monomers require milder deprotection conditions than HF or TFMSA, it does not allow orthogonal synthesis with the commercially available Fmoc-protected amino acids. To carry out orthogonal synthesis of peptide-PNA conjugates, a number of strategies have emerged.[4, 15] For example, the synthesis of Dde/Mmt-protected PNA monomers **1.7** has been described (Figure 1.5).[4, 20] The Dde group was described by Chhabra in 1993 and is selective for primary amine groups.[21] Dde deprotection is typically carried out under nucleophilic conditions (transenamination) with 2% (v/v) hydrazine in DMF, which also deprotects the Fmoc group.[21] To have the Dde group fully orthogonal to the Fmoc group, Bradley developed specific Dde deprotection conditions using a mixture of hydroxylamine hydrochloride and imidazole.[20] These conditions are milder than hydrazine, compatible with many functional groups and resins. Moreover, the Dde group is stable to Fmoc deprotection conditions (solution of 20% piperidine in DMF) and TFA treatments. Other protecting groups (Aloc, Azoc, 4-N_3Cbz) that are orthogonal to Fmoc chemistry have also been used to protect the primary amino group of the backbone.[15] However, Aloc deprotection is often problematic as it needs to be carried out under neutral conditions using palladium catalysts, and it was shown that Aloc/Mmt-protected PNA monomers afforded PNA-peptide conjugates in low yields and purities.[4] Similarly neutral conditions are necessary to remove the Azoc (using phosphines) and 4-N_3Cbz (using phosphines followed by trichloroacetic acid to decompose the azaylide into an amine) groups.[15]

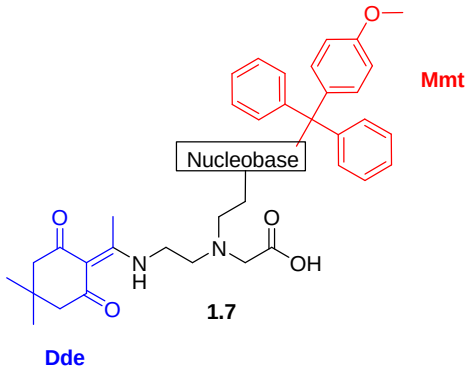


Figure 1.5: Structure of Dde/Mmt-protected PNA monomers (Thymine PNA monomers do not require nucleobase protection)

1.1.4 Applications

Thanks to the enhanced features of PNAs described in section 1.1.2, PNAs have been applied in numerous biological applications. PNAs can be used in a molecular-biology and functional genomics role as well as a diagnosis or detection tool.[2, 22]

1.1.4.1 PNAs in a molecular-biology and functional genomics role

Strand invasion of double-stranded DNA by PNAs have favoured the use of PNAs as antigene and antisense agents.[23] The therapeutic potential of PNAs has been demonstrated by the use of PNAs for example against viral[24] and bacterial[25] infections, and cancers.[26]

In an antigene strategy, transcription of DNA can be inhibited by PNAs, through formation of a strand-invaded triple-helix that induces a sterical hindrance and blocks activity of RNA polymerase.[27] Antigene PNAs, designed to target the transcription-initiation sites for the human progesterone receptors A and B, have been shown to inhibit transcription inside cells by binding to complementary DNA sequences present in the complex normally formed between DNA, RNA polymerase and transcription factors.[28] *In vitro* replication of DNA can also be inhibited by PNAs *via* triple helix formation,[29] and plasmid DNA replication has been blocked *in vivo*.[30]

In an antisense way, PNAs are designed to hybridise to complementary sequences of messenger RNA (mRNA) targets, inhibiting translation of the target genes (*e.g.* through the formation of a triple helix structure that induces a steric blockage).[31] Antisense PNAs have also been reported to inhibit microRNAs.[32] However, an

important issue in the development of antisense experiments is the poor cellular penetration of PNAs.[33] In fact, due to their non-ionic natures, PNAs are often unable to enter into target cells *via* the use of classical transfection agents (*e.g.* cationic lipids and polymers). A number of methods such as electroporation,[34] co-transfection with DNA,[35] and more recently conjugation of PNAs to cell-penetrating peptides[36, 37] (*e.g.* arginine-rich peptides, penetratin) or steroids[38] have been described to enhance cellular uptake of PNAs. For example, fluorescein-labelled PNAs conjugated to a hepta-arginine peptide or the Tat_{48-57} decapeptide were synthesised using Dde/Mmt-protected PNA monomers **1.7** (Figure 1.5) and efficiently delivered into cells.[36] An alternative is the incorporation of positively charged residues (*e.g.* lysines, arginines) into the PNA backbone: recently, it was shown that a guanidine-based PNA (GPNA) (Figure 1.6), a cell-permeable PNA analogue where D-arginines are incorporated into the backbone instead of glycine side chains, blocked protein translation.[39]

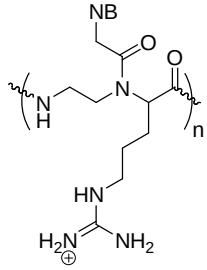
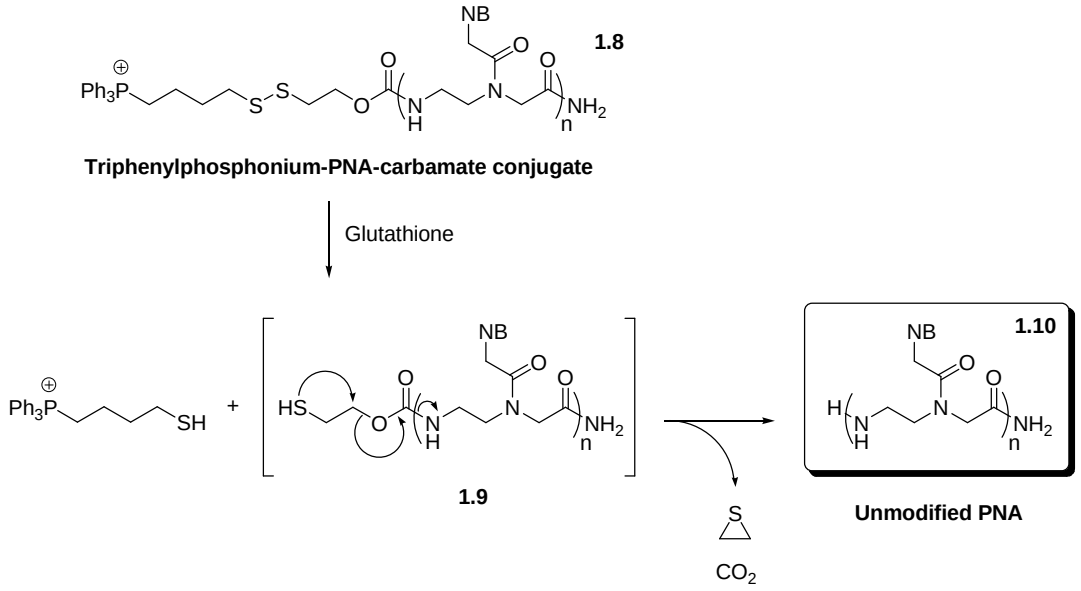


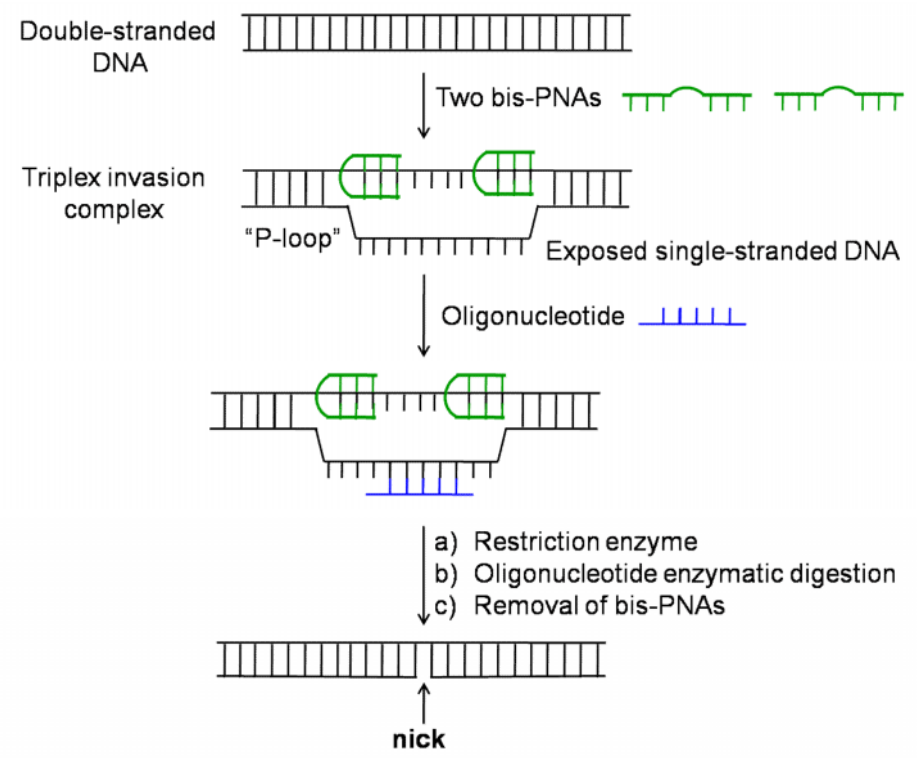
Figure 1.6: Chemical structure of a GPNA, where NB is one of the four natural nucleobases

Another efficient cellular delivery system relies on the introduction of a triphenylphosphonium-PNA-carbamate conjugate **1.8** that is degraded in the cytosol by glutathione into an unstable intermediate **1.9**, leading to the release of an unmodified PNA **1.10** (Scheme 1.1).[40]



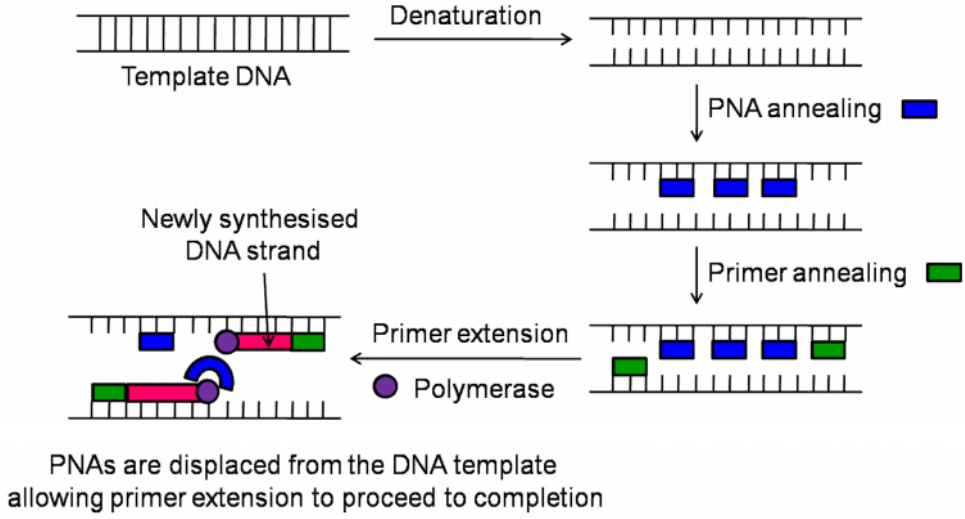
Scheme 1.1: Mechanism of PNA intracellular release from a triphenylphosphonium-PNA-carbamate conjugate

PNAs, used in combination with methylases and restriction enzymes, have been described as a new class of rare genome cutters.[10] A short bis-PNA (Figure 1.4C) is first bound to DNA at a specific site, being consequently protected from enzymatic methylation, while all the other unprotected sites are methylated. Following removal of the bis-PNA, cleavage by the methylase-sensitive restriction enzyme occurs at the designated site, allowing large genomic DNA to be cut into a number of pieces. A PNA-based system has also been developed to design artificial DNA nickases that enable the cleavage of selected sequences in one strand of double-stranded DNA.[11] This system takes advantage of the ability of bis-PNAs to open-up DNA duplexes by invasion (Figure 1.4C), thus exposing a single-stranded DNA that hybridises with an oligonucleotide. The resulting DNA duplex is then cut at a specific site by a restriction enzyme, followed by enzymatic digestion of the oligonucleotide and removal of the bis-PNAs, creating a nick into one strand of the initial DNA duplex (Scheme 1.2).



Scheme 1.2: Design of artificial DNA nickases *via* a PNA-based system

Duplexes formed between a PNA linked to six histidines and a target complementary DNA have been used in combination with nickel affinity chromatography to purify nucleic acids.[41] The efficiency of the polymerase chain reaction (PCR) of variable number tandem repeat commonly used in genetic typing can be enhanced by using PNAs.[42] Small allelic products of heterozygous individuals are preferentially amplified in comparison to large alleles, which can induce false genotyping patterns. To circumvent this problem and obtain reliable PCR amplification of different size alleles, PNAs are used to block the DNA template during amplification (the template is thus unavailable for inter-strand and intra-strand interactions) but not the primer extension during which the polymerase displaces PNAs from the template (Scheme 1.3).



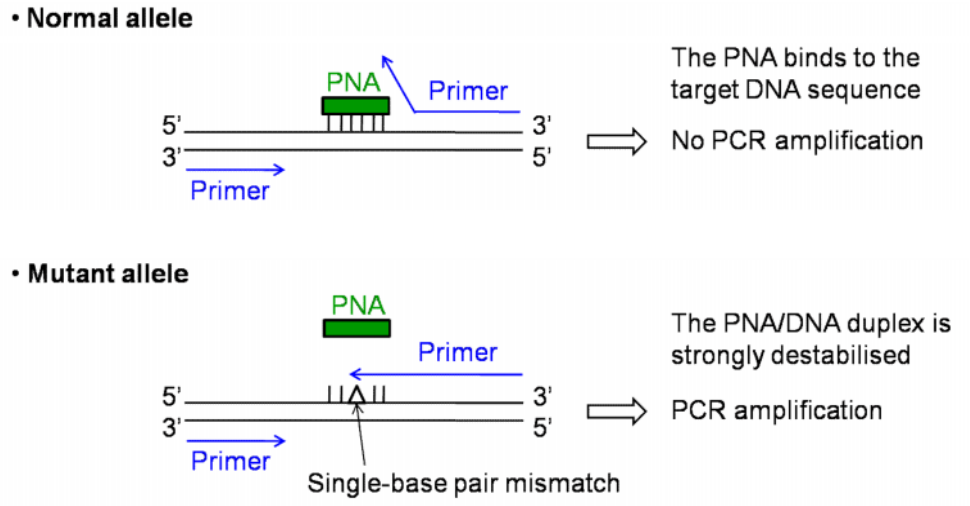
Scheme 1.3: Schematic representation of a modified PCR reaction using PNA[42]

As a functional genomics tool, PNAs have been utilised to improve the southern hybridisation technique that predicts size and sequence information of DNAs.[43] To do so, biotinylated PNAs are first hybridised to denatured DNAs at low salt concentrations (to disfavour formation of DNA/DNA duplexes) followed by size separation of single-stranded DNAs by gel electrophoresis. The PNA/DNA duplexes are then transferred onto a nylon membrane and detected using chemiluminescence-based methods.

1.1.4.2 PNAs as a diagnosis or detection tool

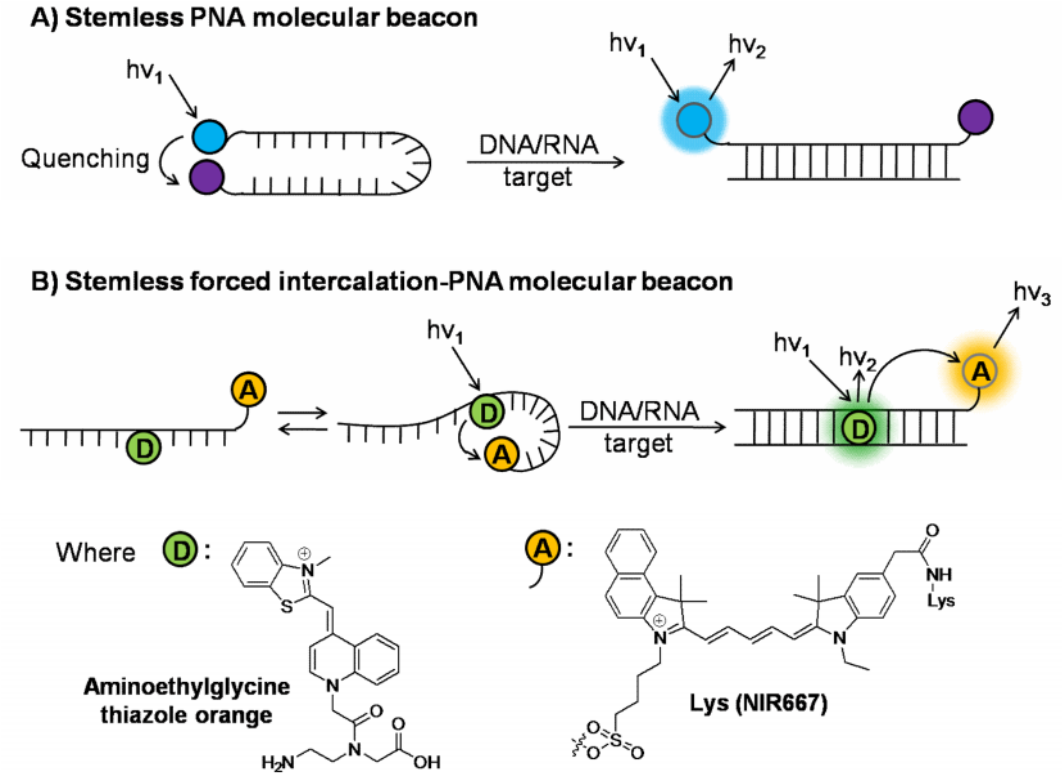
The unique properties of PNAs have allowed the development of diagnostic assays and applications in genetics. For example, PNA-fluorescent *in situ* hybridisation (FISH) technique uses fluorescently labelled PNA probes to visualise nucleic acid targets inside cells.[44] The PNA-FISH method was first applied to determine the accurate telomere length of chromosomes using a fluorescein labelled PNA,[45] and is currently used in cancer and aging studies. Another application is the detection of single base-pair mutations or single-nucleotide polymorphisms (SNPs) in human genetic diseases using PNA-directed PCR clamping method.[46, 47] In the PCR clamping technique, a PNA sequence is designed to bind to one of the PCR primer sites, thus blocking formation of a PCR product (PNA does not act as a primer for DNA polymerase[2]). However, when a single base-pair mismatch is present in the

target DNA sequence, the PNA/DNA duplex is strongly destabilised due to a decrease in its T_m, allowing preferential amplification of the mutated sequence (Scheme 1.4).



Scheme 1.4: Schematic representation of the PNA-directed PCR clamping method[46]

PNAs that are labelled with a fluorescent dye and a quencher at each terminus, also known as molecular beacons, have been used as biosensors for nucleic acid detection. When not hybridised, the donor and the acceptor labels are in close proximity so that the fluorescence is quenched. Upon binding to a complementary DNA or RNA sequence, the fluorophore is distanced from the quencher, leading to an increase in the dye fluorescence (Scheme 1.5A).[48] Recently, this concept has been adapted to enhance fluorescence and have a better mismatch discrimination, even at low temperatures. Stemless forced intercalation-PNA molecular beacons have been designed to have an open conformation containing an intercalator dye (*e.g.* thiazole orange) used as a donor and at one end a near infrared dye (*e.g.* NIR667) used as an acceptor. Upon hybridisation, the intercalator dye becomes fluorescent if enclosed within perfectly matched base pairs, and its fluorescence is subsequently quenched by the near-infrared dye (Scheme 1.5B).[49]



Scheme 1.5: Detection of complementary target nucleic acids using A) PNA molecular beacons[48] and B) stemless forced intercalation-PNA molecular beacons (adapted from Socher[49])

1.2 Methods of solid phase synthesis

1.2.1 Introduction

Traditional organic reactions are carried out in solution phase to produce one compound isolated after a work-up and purification. To avoid these time-consuming steps and solubility issues associated with peptide synthesis, Merrifield introduced in 1963 the concept of solid phase chemistry.[50] In solid phase synthesis, molecules are attached to a functionalised resin containing a linker. Reagents can be added in excess without complicating the later purification of the product, which is done by simple washings of the resin with solvents. Solid phase techniques have been developed to allow parallel and split and mix syntheses of a range of compounds.

1.2.2 Multipin method

The origins of parallel synthesis can be traced to 1984, when Geysen described the multipin method for antibody epitope mapping. Hundreds of individual peptides

were synthesised on polyacrylic acid-grafted polyethylene pins arrayed in a microtitre plate format.[51] Reagents were distributed into the wells of a 96-microtitre plate so that 96 reactions were run in parallel. Common steps of washing, neutralisation and coupling were simply carried out by simultaneously immersing all the pins in specific solutions. Other graft polymers such as polyacrylamide and polystyrene could also be used, and polymers could be functionalised with the classical linkers used in solid phase chemistry. Although the multipin method is an effective and inexpensive technique where pins can be easily manipulated, low loadings are obtained in comparison with standard solid supports.

1.2.3 Tea-bags

Like many other combinatorial methodologies, the tea-tag method was first designed for peptide synthesis, although it can be generalised to other organic reactions. In 1985, Houghten developed the tea-bag method,[52] and reported the synthesis of 248 different peptides obtained in 10-20 mg quantities. Small batches of functionalised resin beads were sealed in polypropylene mesh packets that looked like tea-bags. The advantage of the method was the possibility to combine all the common washing, deprotection and coupling steps, while different amino acid couplings could be performed by immersing each individual packet in a separate solution containing a specific amino acid. At the end of the synthesis, final compounds were cleaved from the solid support, identified *via* the label placed on each tea-bag, and obtained in yields and purities comparable to traditional techniques. This evolved into other techniques such as Irori resin carriers (Figure 1.7). Libraries of compounds are synthesised inside these polypropylene microreactors using a radiofrequency[53] (*e.g.* MacroKans, MicroKans) or optical[54] (*e.g.* NanoKans) tagging system.

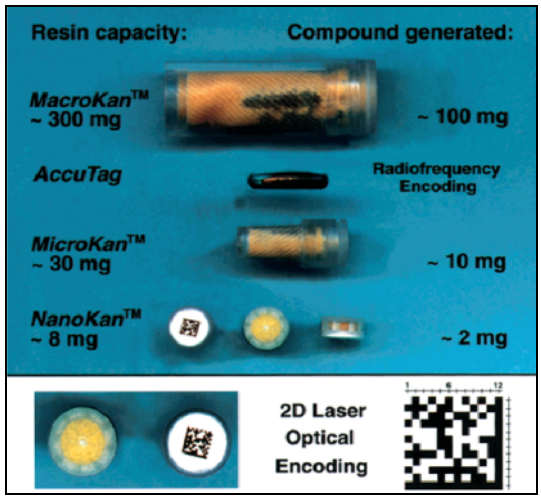


Figure 1.7: Different types of Irori resin carriers (reproduced with permission)[54]

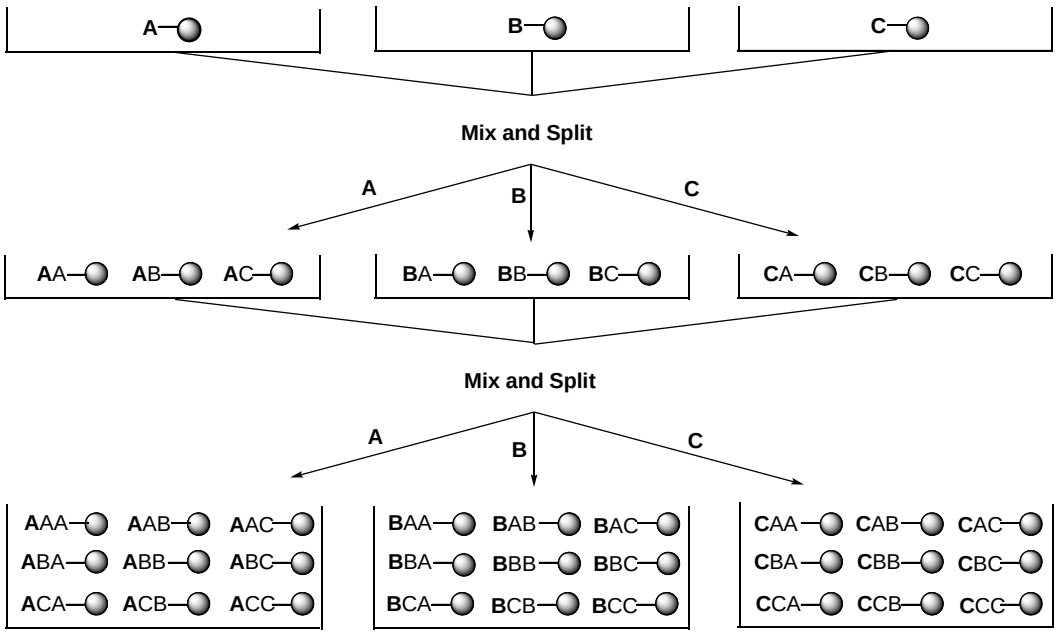
1.2.4 SPOT synthesis

SPOT synthesis, developed by Frank in 1992, allowed the rapid and cost-effective simultaneous parallel synthesis of large numbers of peptides arrayed on a cellulose membrane.[55] To carry out peptide synthesis, membranes were first derivatised with an amino functional group by esterification of the hydroxyl groups on cellulose with Fmoc-β-Ala-OH. In the second step, an amino acid was coupled at a specific location by manual or automatic dispensing of a solution containing the activated Fmoc-protected amino acid in a non-volatile solvent, followed by acetylation of the remaining reactive amino groups. If required, peptides can be released from the support by introducing a cleavable linker on the derivatised membrane. SPOT synthesis, which uses simple protocols and small amounts of reagents, has been applied in many biological applications such as screening for antibody binding or determination of enzyme substrate specificities.

1.2.5 Split and mix method

The split and mix method was reported by Furka in 1988 for the synthesis of libraries of equimolar peptide mixtures.[56, 57] In this technique, the beads are divided into separate groups equal to the number of different introduced building blocks. Following couplings of the reagents to each batch of beads, the entire set of beads is mixed in one pool and split again for the next coupling. So the synthesis of a library

with three building blocks (A, B and C) used in three steps yields 27 (3^3) products (Scheme 1.6). The library size increases with the number of steps so that a tetrapeptide library with 20 natural amino acids leads to the generation of 160,000 (20^4) different compounds, while requiring only 80 coupling reactions! This powerful methodology avoids reactions of competition between introduced reagents, and products are present in equimolar quantities in the final mixture. Moreover, one bead interacts with only one reagent so that each bead displays one type of compound, hence the concept of "one-bead-one-compound" described by Lam in 1991.[58] However, since the library is obtained as a mixture, deconvolution and encoding strategies are required to identify individual products.



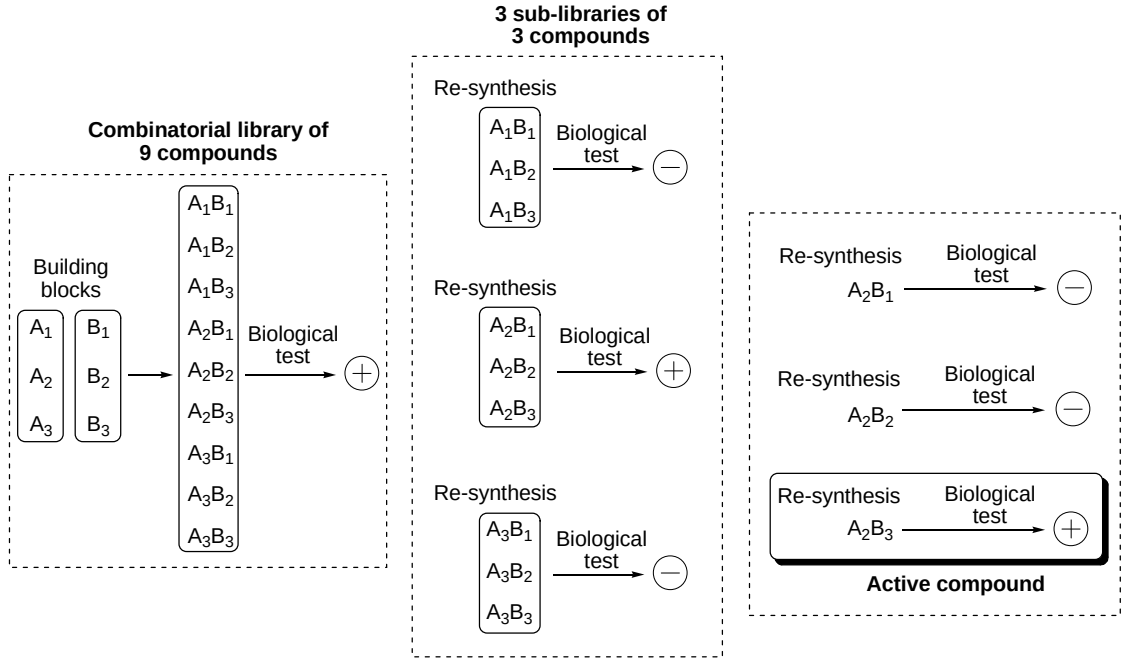
Scheme 1.6: Principle of the split and mix method

1.3 Deconvolution strategies

Different deconvolution approaches have been described for the identification of biologically active single compounds from mixtures in a combinatorial library. Two strategies (iterative deconvolution and positional scanning) are described below.

1.3.1 Iterative deconvolution

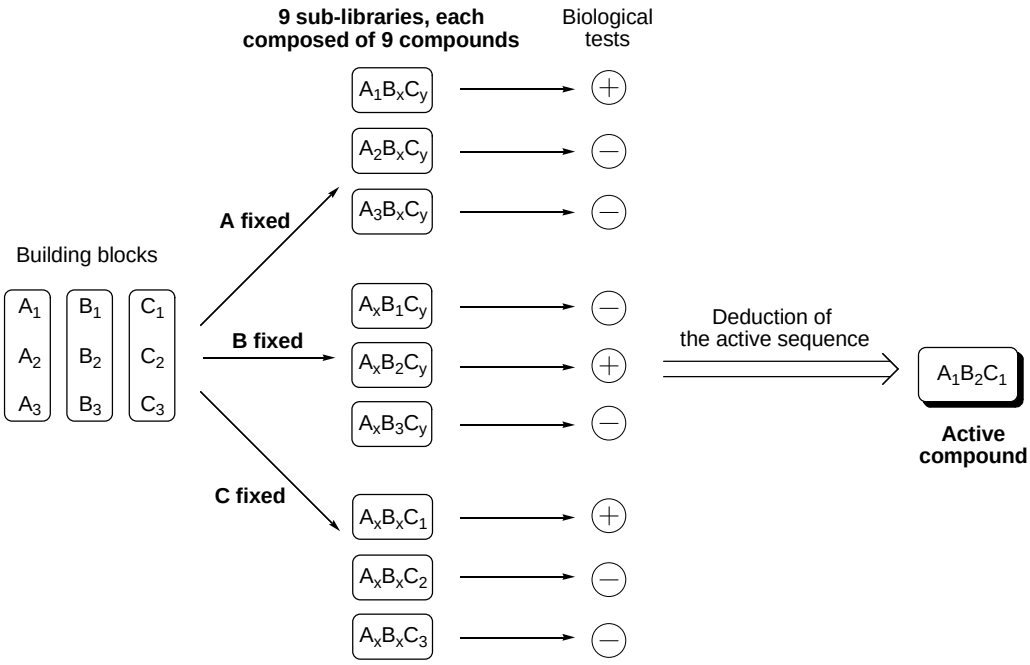
The iterative approach was introduced by Geysen to deconvolute pin-immobilised peptide mixtures,[59] and generalised by Houghten for deconvolution of soluble mixture-based combinatorial libraries.[60] It involves successive steps of synthesis, screening and selection of the most active mixtures. For example, consider a combinatorial library synthesised using a split and mix methodology and composed of 9 compounds obtained in two successive steps (A_1, A_2, A_3 are used as building blocks for the first step and B_1, B_2, B_3 for the second one). A biological test is first carried out. If a biological activity is detected (positive test) then three sub-libraries containing one fixed building block (A_1, A_2 or A_3) are re-synthesised by separate reactions of A_1, A_2 and A_3 with a mixture of B_1, B_2 and B_3. A biological test is carried out on each sub-library. Compounds of the sub-library that give a positive test are individually re-synthesised and screened for identification of the active product (Scheme 1.7). However, an important drawback of the iterative concept is that it is time-consuming due to the requirement for re-syntheses and repetitive biological screenings.



Scheme 1.7: Principle of iterative deconvolution

1.3.2 Deconvolution by positional scanning

An alternative to the iterative approach is deconvolution by positional scanning, a concept developed by Houghten in 1992 that allowed identification of key compounds at each library position.[61] A positional scanning library is divided into a number of separate sub-libraries equal to the number of introduced reagents. In each sub-library, one position is fixed with a single residue, while the other positions contain a mixture of building blocks. All sub-libraries include the same variety of products with a unique difference the location of the fixed position. Each sub-library is then screened to instantly determine the best building block at a defined position, and by combining the positive results, the sequence of the most biologically active compound is identified (Scheme 1.8). This process can be used to generate consensus sequences. Unlike the iterative process, no steps of re-synthesis are required and it is faster to determine active individual sequences.

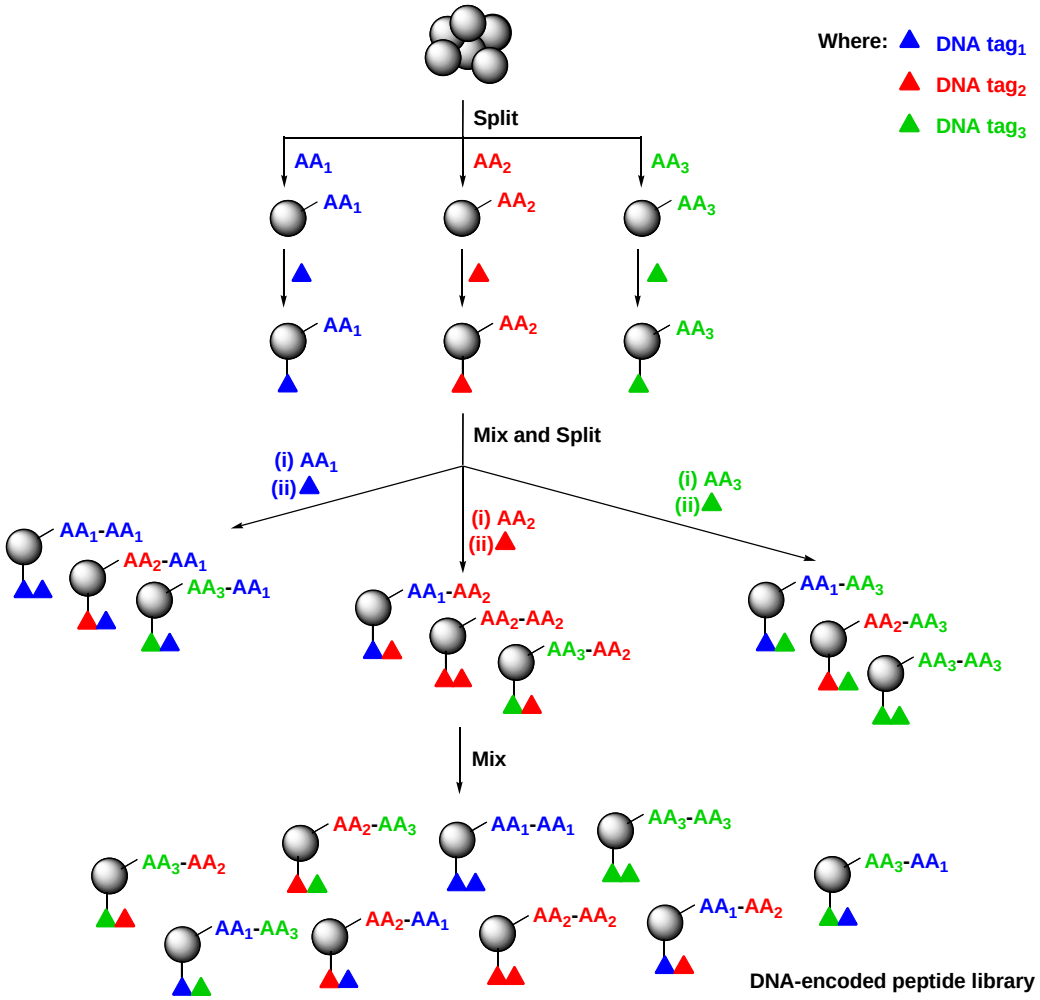


Scheme 1.8: Principle of deconvolution by positional scanning

1.4 Encoded combinatorial libraries

1.4.1 Chemical encoding

In 1992, Brenner introduced the theoretical principle of DNA-encoded combinatorial peptide libraries that could be prepared by alternating synthesis of amino acids and DNAs on the same solid support using a split and mix strategy (Scheme 1.9).[62] Each peptide would thus be encoded by a unique DNA tag. Following affinity capture on a target protein, the DNA tags of the selected compounds would be amplified by PCR and sequenced to reveal the identity of the peptide sequences. In 1993, Affymax[63] and Nielsen[64] independently demonstrated the validity of this concept. In both methods, two orthogonal groups (the base-labile Fmoc group and the acid-labile dimethoxytrityl (DMT) group) were used to allow orthogonal synthesis of peptides and DNAs. Affymax reported the synthesis of a DNA-encoded library composed of 823,543 heptapeptides.[63] The solid support (polystyrene beads) was functionalised with a mixture of Fmoc-Thr(tBu)-OBt and the succinimidyl activated ester of 4-*O*-DMT-oxybutyrate that were used to carry out peptide and DNA synthesis respectively (Figure 1.8). Using a spectrophotometer, it was found that the construct **1.11** contained a 20:1 ratio of Fmoc-protected amino groups to DMT-protected hydroxyl groups. Nielsen employed a different construct **1.12** that contained a serine residue as a core. Controlled-pore glass beads were chosen as the solid support, the pentapeptide leucine enkephalin was synthesised *via* the α-amino group of serine and its DNA tag was built *via* the hydroxyl group of the serine side chain (Figure 1.8). In this strategy, upon cleavage of the peptide from the solid support, the peptide remained linked to its encoding DNA tag in a 1:1 ratio.[64]



Scheme 1.9: Schematic representation of the DNA-encoded peptide library concept proposed by Brenner[62]

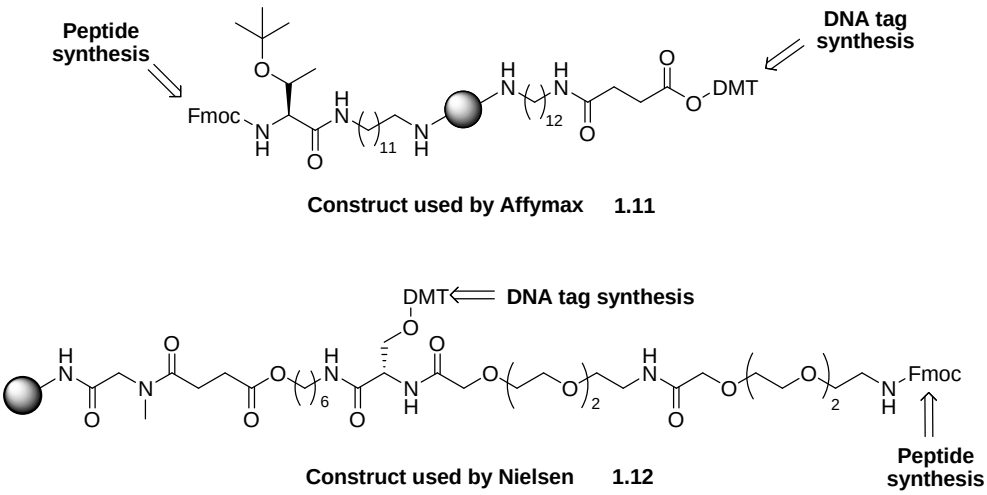
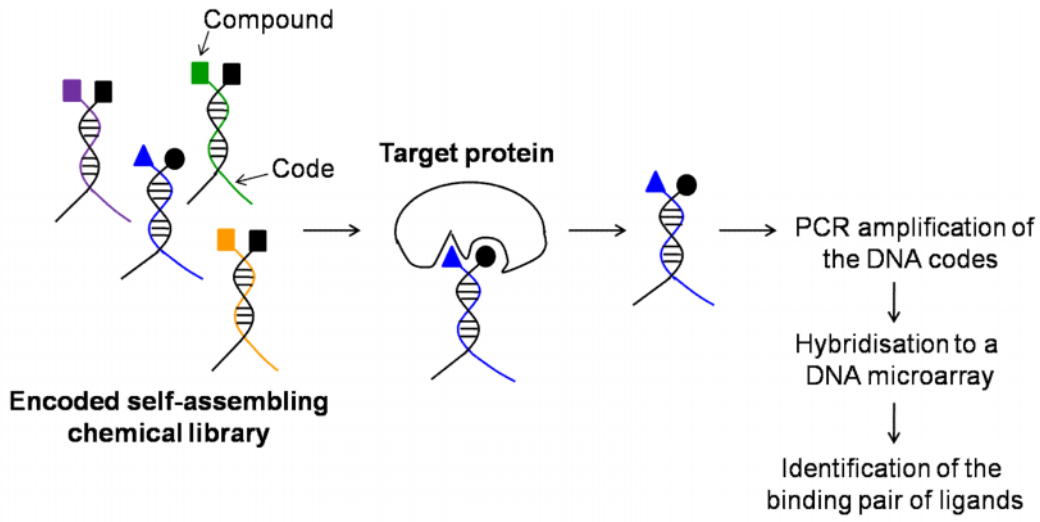


Figure 1.8: General structures of the constructs used by Affymax and Nielsen for DNA-encoded peptide library synthesis

Several DNA-encoded libraries were constructed.[65] Recently, encoded self-assembling chemical (ESAC) libraries have been reported as an efficient tool to identify molecules that bind target proteins.[66] In this technique, separate sub-libraries of oligonucleotide-compound conjugates are synthesised, with each compound encoded by a specific DNA code. By mixing two sub-libraries, complementary DNAs can hybridise together to form a DNA duplex. As a result, a very large ESAC library with diverse chemical groups is generated and incubated with a target protein. Molecules that bind to the protein with high affinity are eluted, while the unbound molecules are removed by washing steps. The unique DNA encoding codes of the binders are amplified by PCR and hybridised to a DNA microarray to "decode" and identify the structure of the selected pair of ligands (Scheme 1.10). Another example of DNA-encoded libraries involving enzymatic DNA ligation has been patented by the companies Nuevolution[67] and Praecis.[68] A first building block (*e.g.* amino acid, small molecule) is chemically linked to an encoding DNA that contains a reactive functional group, and a second building block is introduced to react with the first one. A DNA that encodes this second building block is then enzymatically ligated (*e.g.* using DNA ligase) to the first DNA. The procedure is repeated using a split and mix methodology to end up with a DNA-encoded library that can contain more than 10^9 compounds.



Scheme 1.10: Schematic representation of the selection of a binding molecule from an ESAC library

Still developed inert haloaromatic tags **1.13** that attached a tag to the resin bead (Figure 1.9). Following bead screening, bead removal, tag cleavage and derivatisation, the tags could be detected at the sub-picomolar scale by electron-capture gas chromatography.[69]

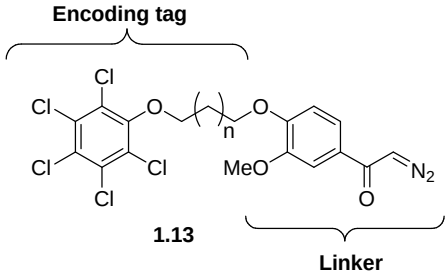
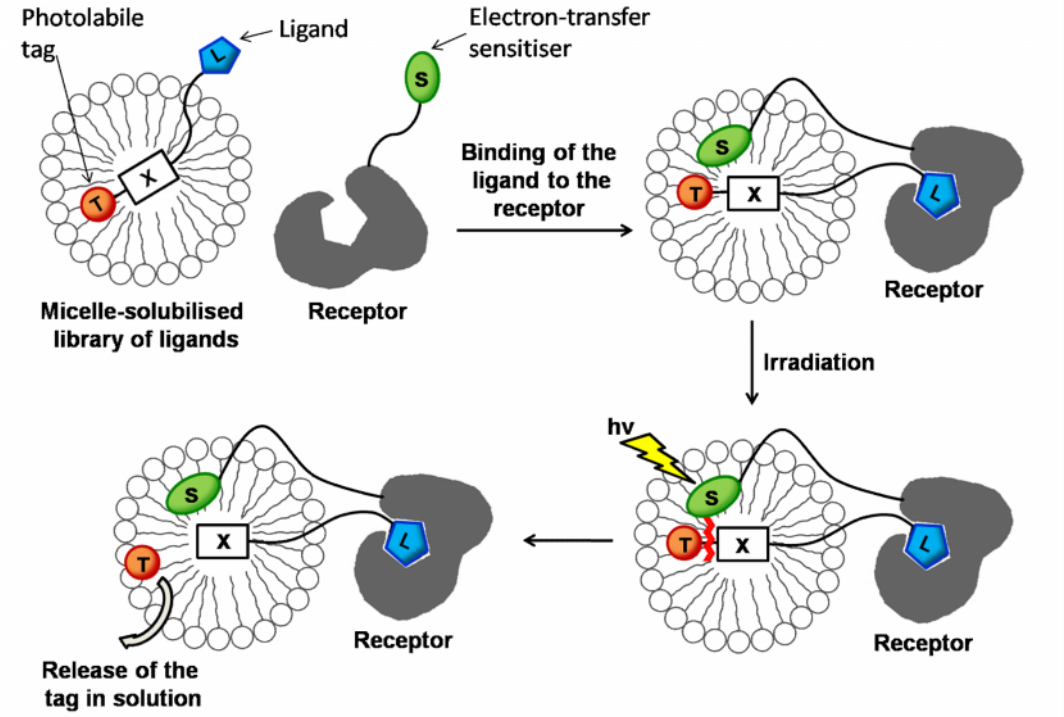


Figure 1.9: General structure of a haloaromatic tag

Another example of a chemical encoding method is the use of sensitised photolabile tags.[70] A 2-alkyl substituted dithiane-based photocleavable molecule is attached to a ligand while an electron-transfer sensitiser is linked to a receptor. Upon binding of the ligand to the receptor, the sensitiser is brought into close proximity with the photolabile tag, and following irradiation, the tag is released in solution, allowing identification of the active ligand (Scheme 1.11). Peptides, secondary amines and fluorescent dyes,[71] as well as ^{19}F NMR and isotopic labels[72] have also been used as tagging techniques.[73]



Scheme 1.11: Use of sensitised photolabile tags to encode solution phase combinatorial libraries (adapted from Kottani)[70]

1.4.2 Non-chemical encoding

As chemical tags can interfere with the versatility of chemical synthesis, a number of non-chemical encoding methods (*e.g.* spectrometric, electronic, graphical and physical encoding) have been described.[74] To encode combinatorial libraries (using a tea-bag approach), Nicolaou developed a radiofrequency tagging system, where semiconductor memory microchips were enclosed within a porous and inert microreactor along with resin beads.[53] At each reaction step, the microreactor was scanned and a unique radiofrequency signal was recorded on every microchip so that the synthetic path of each bead was known and the identity of the product could be deduced. Large peptide combinatorial libraries (at the single bead level) have been encoded with different fluorescent colloids (containing one, two or three fluorescent dyes) that are attached non-covalently to the surface of resin beads during each amino acid coupling reaction. Each resin bead is thus composed of specific fluorophore layers that produce a fluorescent "barcode" decoded using fluorescence microscopy (Figure 1.10).[75]

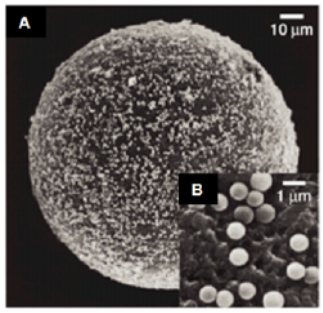
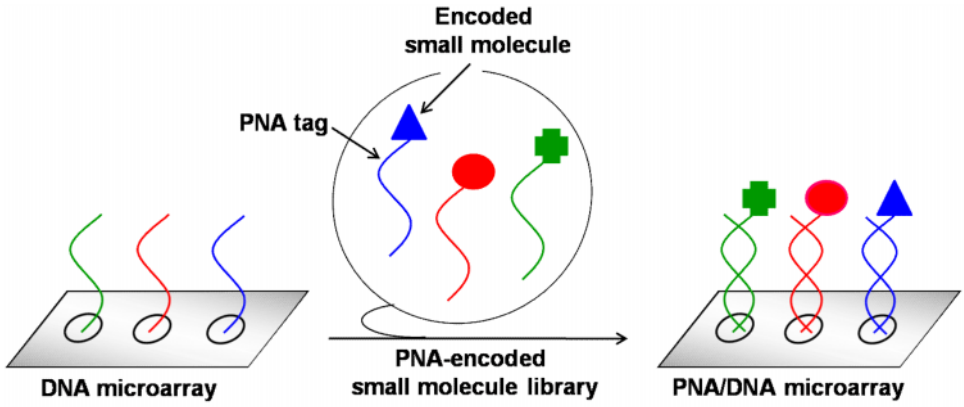


Figure 1.10: Scanning electron microscope images of fluorescent colloidal particles (B) non-covalently attached to the surface of a polystyrene resin bead (A) (reproduced with permission)[75]

1.4.3 PNA-encoded libraries

Unlike DNA that is sensitive to harsh conditions of organic synthesis, PNA is extremely resistant, opening the door to new chemical encoding applications.[76] PNA has been used as a tagging molecule,[77] with PNA-encoded libraries being successfully used to study enzyme substrate specificity.[78-82] Winssinger synthesised PNA-encoded small molecule libraries, in which each PNA tag was used to encode a small molecule and allow identification of its structure by hybridisation to a predefined location on a DNA microarray (*via* formation of a highly stable PNA/DNA duplex) (Scheme 1.12).[78, 79]



Scheme 1.12: Hybridisation of a PNA-encoded small molecule library on a DNA microarray

In proof-of-concept experiments, it was shown that the PNA tags did not alter the activity of the small molecules against a protease.[78] PNA-encoded libraries have been used to identify protease inhibitors.[79, 80] Six fluorescein-labelled PNA-encoded

cysteine protease inhibitors (Figure 1.11) were synthesised on Rink amide resin. Lysine was first attached to the resin and protected with the Fmoc and Aloc orthogonal protecting groups to allow synthesis of the peptide acrylate-based inhibitor and PNA tag respectively. The library was incubated in solution with a target protease (cathepsin) and the PNA-encoded inhibitors that bound to the protease were separated from the unbound ones by size exclusion filtration (or using a gel-based method[80]) and hybridised to a DNA microarray to decode their structures *via* the PNA tags. The approach was also used to measure cysteine protease activity in complex mixtures of proteins (*e.g.* crude cell lysates) and identify human caspase inhibitors.[79]

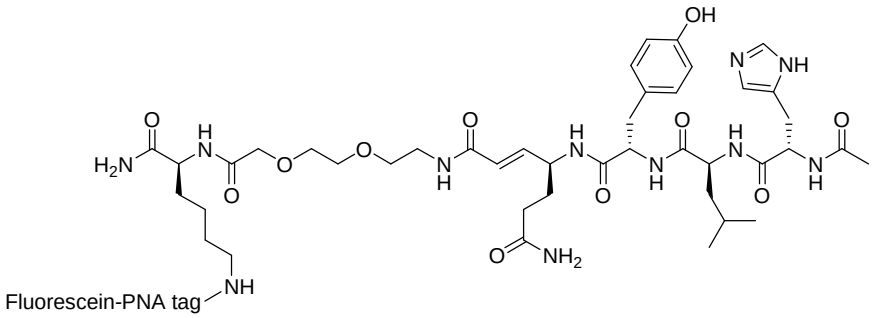


Figure 1.11: Structure of a PNA-encoded cathepsin L inhibitor

Bradley reported the syntheses of PNA-encoded split-and-mix peptide libraries to determine the substrate specificity of different proteases and the abelson tyrosine kinase.[81, 82] In this strategy, the libraries were obtained through alternative couplings of Fmoc-protected amino acids and Dde/Mmt-protected PNA monomers, which was made possible thanks to the development of orthogonal deprotection conditions between the Fmoc and Dde protecting groups (see section 1.1.3).[20] Fmoc-Lys(Dde)-OH was chosen as a core to synthesise the peptide-PNA conjugates. Amino acids were coupled on the *N*-terminal side of Fmoc-Lys(Dde)-Rink-PEGA resin **1.14** using an Fmoc strategy, while the PNA tag was built up on the *N*-terminal side chain of lysine using Dde chemistry (Figure 1.12). For instance, a 10,000-member PNA-encoded peptide library was used to profile chymopapain and subtilisin by using fluorescein and tetramethyl-6-carboxyrhodamine (TAMRA) as a fluorescence resonance energy transfer (FRET) system (Figure 1.13).[82] Each amino acid was

encoded by a PNA triplet so that the peptide-PNA conjugates could be incubated with the target protease in solution (it represented an advantage over solid support screening approaches since the interaction between the peptide and the protease was more similar to physiological conditions and it avoided unspecific reactions between the peptides and the solid surface) and then hybridised to a complementary DNA microarray to identify the peptide sequences recognised by the protease.

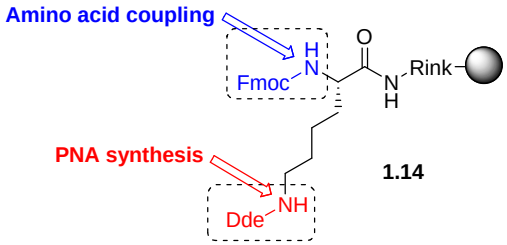


Figure 1.12: Fmoc-Lys(Dde)-Rink-PEGA resin as a core to prepare PNA-encoded peptide libraries

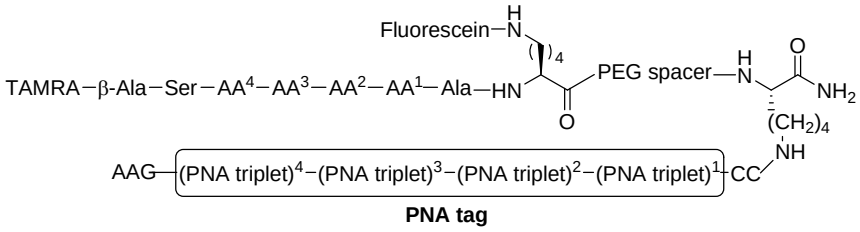


Figure 1.13: General structure of the PNA-encoded FRET-based peptide library

1.5 Microarray technology

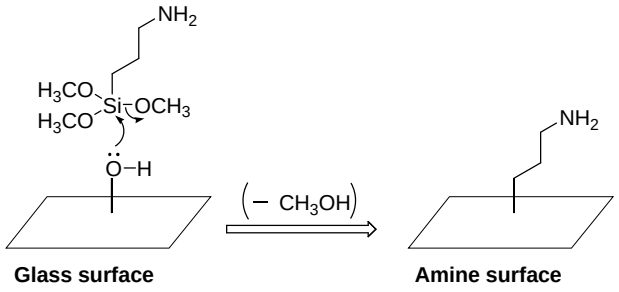
1.5.1 Introduction

Several high-throughput methods have been developed to rapidly generate a large number of compounds. One key feature is to identify active compounds against hundreds of biological targets using rapid and inexpensive technologies. Recent improvements such as miniaturisation and automation of biological assays have favoured the expansion of microarray technology.[83] It is a powerful methodology that allows the simultaneous study of thousands of molecules in a single experiment. Indeed a microarray consists of large numbers of known biological molecules arranged in a structured grid like format. In a typical experiment, biomolecules are

immobilised onto a solid support (*e.g.* a glass slide or membrane) at defined positions, followed by a specific reaction with labelled targets. The resulting fluorescent signal is analysed to indicate the identity and the abundance of each labelled target in the starting mixture. Significant advantages are the use of tiny reaction volumes (microliter volumes) and amount of samples (at the nanomole- and picomolescale) to carry out microarray experiments. Microarray technology has been applied to the high-throughput analysis of different biomolecules. A variety of formats have been developed including DNA,[84] protein,[85] peptide,[86] tissue,[87] carbohydrate,[88] live cell,[89] small molecule[90] and polymer[91] microarrays. It represents a powerful tool in many research areas such as genetic screening, proteomics, drug discovery and diagnostics.[83]

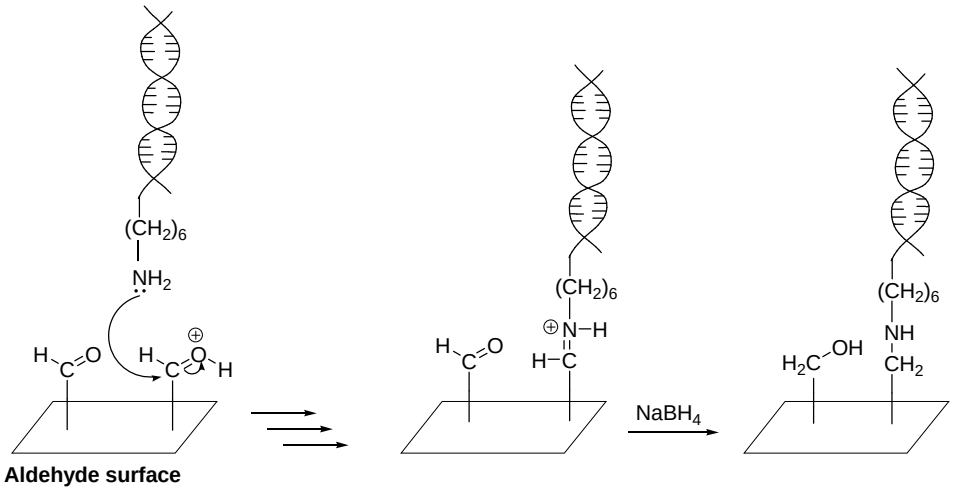
1.5.2 Types of surfaces and immobilisation

The choice of a microarray surface represents an important criterion. High-quality microarray surfaces are required to ensure efficient immobilisation of biomolecules and to avoid unspecific binding with labelled targets. The surface can also influence the reaction efficiency, the quality of the final data, and it needs to be compatible with the detection method, for example a low background if used in conjunction with fluorescence. The most commonly used material is glass, whose main component is silicon dioxide. Glass slides are flat, planar, transparent, resistant to high temperatures, easy to handle (standard dimensions of a glass microscope slide: 75 mm x 25 mm), and a key point is the possibility of modifying their surfaces.[92] Traditional amine, aldehyde and epoxide surfaces have been developed and are commercially available. An amine-derivatised surface is created by reacting a glass surface with 3-aminopropyltrimethoxysilane (Scheme 1.13) (aldehyde and epoxide surfaces are prepared in a similar manner).[83]



Scheme 1.13: Covalent attachment of a reactive primary amine to a glass surface[92]

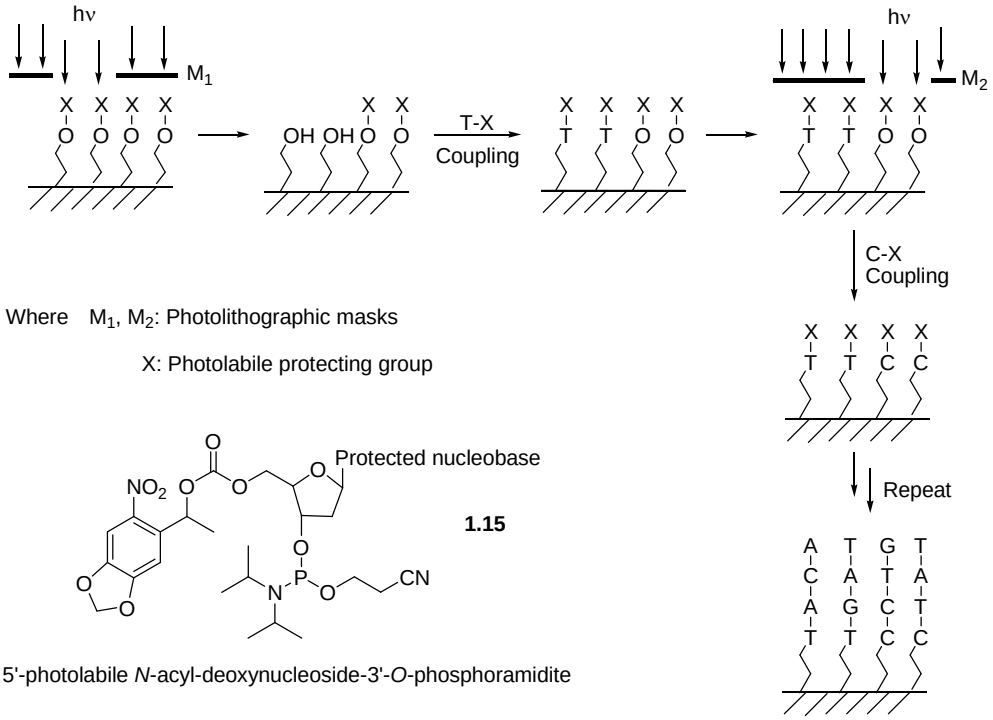
Biomolecules can be immobilised on microarrays either *via* adsorption consisting of non-covalent interactions between the biomolecules and the slide surface (*e.g.* DNAs immobilised on an amine surface through electrostatic interactions between the phosphate groups of DNAs (negatively charged) and the positively charged amino groups) or *via* the formation of a covalent bond. For instance, the attachment of an amino-modified DNA to an aldehyde surface forms a covalent imine bond that is reduced into a stable secondary amine using sodium borohydride. This treatment also minimises the fluorescent background by reducing unreacted aldehyde groups into non-reactive primary alcohols (Scheme 1.14).[92]



Scheme 1.14: Immobilisation of an amino-modified DNA on an aldehyde surface[92]

1.5.3 DNA microarrays

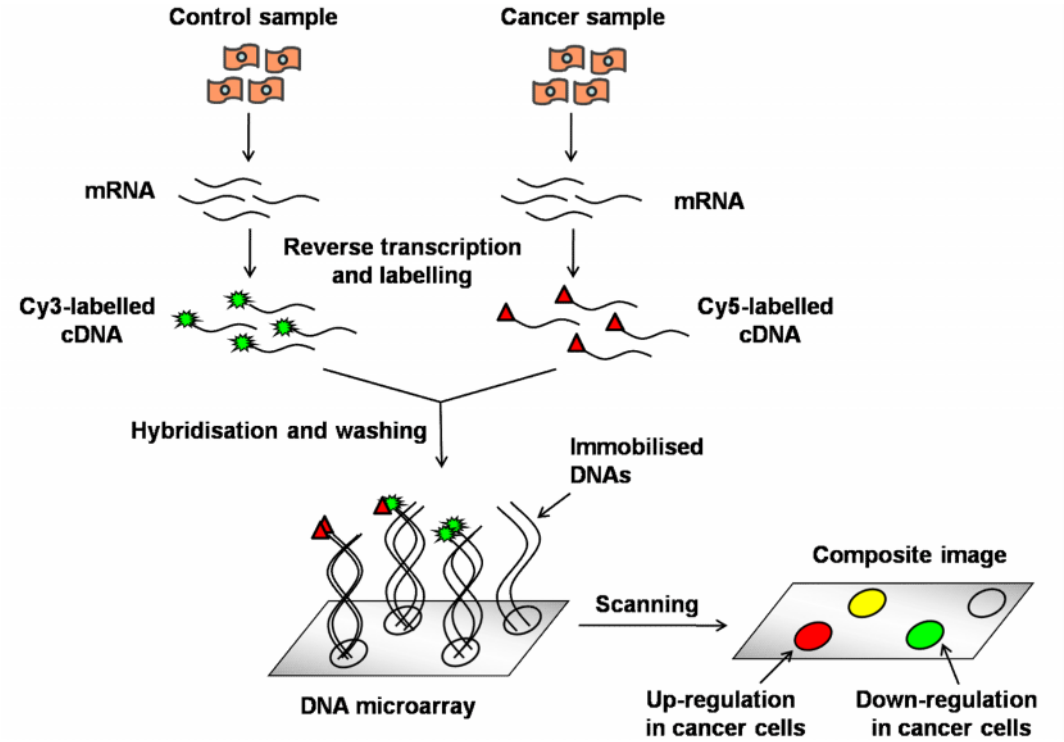
In 1994 Fodor from Affymax synthesised an array of 256 different tetranucleotides using photolithography.[93] In this technique, a photolithographic mask was applied to predefined regions of a glass surface containing photolabile-protected hydroxyl groups. Following exposure to light, the photolabile protecting groups were selectively removed in the regions not protected by the mask. A 5'-photolabile *N*-acyl-deoxynucleoside-3'-*O*-phosphoramidite **1.15** was then added to the entire surface to react with the free hydroxyl groups. Following capping, oxidation and washing steps, a second photolithography mask was applied to other regions, and after exposure to light, a second oligonucleotide with a photolabile 5'-protecting group was coupled. The process was repeated until the expected oligonucleotide sequences were synthesised *in situ* on the glass surface (Scheme 1.15).



Scheme 1.15: Preparation of a DNA microarray using photolithography (adapted from Pease[93])

In 1995 Schena used a complementary DNA (cDNA) microarray to quantitatively measure the expression of 45 *Arabidopsis* (small mustard plant) genes in parallel.[84] DNA microarrays are prepared either by spotting pre-synthesised DNAs onto

modified glass slides (*e.g.* aldehyde-derivatised surface, see section 1.5.2), or by carrying out *in-situ* DNA synthesis on the solid surface using photolithography[93, 94] or inkjet printing.[95] The main application of DNA microarrays is gene expression profiling, where the relative abundance of a gene in two different samples (*e.g.* normal and cancer samples) can be identified, with mRNA extracted from each sample and reverse-transcribed into labelled cDNA with two different fluorescent dyes (*e.g.* Cy3 and Cy5). The resulting cDNAs are then mixed in equal amounts and hybridised to a DNA microarray. Following washing and scanning using two different channels, the fluorescent signal for each dye is quantified at each spot location, and the ratio of the fluorescent intensities for the two dyes reveals the level of expression of each gene in the tested sample (Scheme 1.16).[96] Using this approach, it is possible to rapidly identify which specific genes are expressed at high levels in diseases such as cancers so that more targeted treatments can be designed. DNA microarrays can also be used for genome-wide genotyping, resequencing applications, and detection of mutations or SNPs, identifying if an individual is susceptible to develop a specific disease.[97] Another application referred to as array-based "comparative genomic hybridisation" has been reported, where two genomes are compared in a single experiment to determine if particular genes have been amplified or lost. This type of analysis is particularly useful in cancer research to follow the progression of the disease.[98]



Scheme 1.16: Gene expression profiling using a dual-channel DNA microarray

1.5.4 Peptide microarrays

Following the success of DNA microarrays, new studies have focused on the development of other types of formats such as peptide microarrays.[99, 100] The concept of peptide microarrays was first introduced by Geysen who described the multipin method (see section 1.2.2) and later on by Frank using the SPOT synthesis (see section 1.2.4). In fact, the first microarray was not a DNA but a peptide microarray composed of 1024 peptides that was synthesised by photolithography and reported by Fodor in 1991.[101] Similarly to DNA microarrays, peptide microarrays can be synthesised in two main ways, either by *in-situ* synthesis directly on the solid surface (using photolithography or SPOT synthesis) or by immobilising pre-synthesised peptide derivatives (*e.g.* cysteine-, cyclopentadiene- and alkoxyamine-terminated peptides, biotinylated peptides),[99] using a contact arrayer or an inkjet printing system (the latter method is preferred for the synthesis of high-density peptide microarrays). Peptide microarrays have been successfully applied for the high-throughput screening (HTS) of various enzymes including kinases,[85, 86, 102] phosphatases[103, 104]

and proteases,[82, 105] allowing a unique insight into the substrate specificity of these enzymes and the identification of biologically active peptides. Peptide microarrays have also been used in molecular immunology (*e.g.* antibody epitope identification,[106] antibody diagnostics[107]) and peptide interactions with other molecules such as proteins have started to be investigated.[100] In comparison with DNA microarrays, peptide microarrays are less accessible so that a smaller number of applications have been reported. That's why many efforts are focusing on the development of affordable, high quality and high density peptide microarrays that could be used as a routine tool in drug discovery, diagnostic applications, and the discovery of new protein binders and inhibitors.

Chapter 2: A 10,000-member PNA-encoded peptide library for profiling kinases

2.1 Protein kinases

Protein kinases represent one of the largest families of eukaryotic proteins, with more than 500 kinases identified within the human genome.[108] Their basic function is to regulate protein phosphorylation by catalysing the transfer of the γ-phosphate of adenosine triphosphate (ATP) to phenolic groups on tyrosine (so-called tyrosine kinases) or alcohol groups on serine and threonine (so-called serine/threonine kinases). Another class of kinases named "dual specificity" kinases are able to phosphorylate both tyrosine and serine/threonine residues,[108] and some kinases that are mainly found in prokaryotes can phosphorylate histidine residues.[109] Protein phosphorylation is a reversible process, with dephosphorylation controlled by another family of enzymes called phosphatases.[110] Human protein kinases can also be divided into two main groups, the eukaryotic protein kinases (ePKs) and the atypical protein kinases (aPKs). The ePKs (478 human protein kinases identified) contain a similar catalytic domain with conserved sub-domains and common amino acid sequences.[111] The ePKs can be classified into nine groups: (1) the AGC group (including the cyclic-adenosine monophosphate-dependent protein kinase (PKA), cyclic-guanosine monophosphate-dependent protein kinase (PKG) and protein kinase C (PKC) families); (2) the CAMK group (*e.g.* calcium/calmodulin-dependent protein kinase); (3) the CK1 group (*e.g.* casein kinase 1); (4) the CMGC group (including the cyclin-dependent kinase (CDK), mitogen-activated protein kinase (MAPK), glycogen synthase kinase 3 (GSK3) and Cdc2-like kinase (CLK) families); (5) the receptor guanylate cyclase (RGC) group; (6) the STE group (*e.g.* yeast sterile-phenotype kinases); (7) the tyrosine kinase (TK) group [*e.g.* Abelson tyrosine kinase (Abl)]; (8) the tyrosine kinase-like (TKL) group and (9) the "other" group (including ePKs that do not belong to any of the above categories).[108] The aPKs (40 human protein kinases identified) are known to have protein kinase activity but have insignificant sequence similarity to the ePK catalytic domain.[108] Most human protein

kinases phosphorylate at serine or threonine with only about 90 believed to be protein tyrosine kinases, although to date only a small proportion of these have been fully characterised.[112] Protein kinases control many cellular pathways, in particular cellular signal transduction processes such as cell division, differentiation, apoptosis and transcription. Any dysfunction can thus be catastrophic and lead to a variety of disease states such as cancer, cardiovascular disease, inflammation, diabetes, arthritis and neurological disorders (*e.g.* Alzheimer's disease).[113]

2.2 Protein kinase inhibitors

In large pharmaceutical companies, about 25% of research projects are related to protein kinases, focusing on the development of potent protein kinase inhibitors.[114] Humanised antibodies as well as many small-molecule kinase inhibitors have emerged[115] and a number have been approved for clinical use such as fasudil **2.1** for the inhibition of Rho-dependent kinase (treatment of cerebral vasospasm),[116] sirolimus **2.2** for the inhibition of the target of rapamycin (immunosuppressive and antiproliferative properties),[117] imatinib **2.3** for the inhibition of the breakpoint cluster region (Bcr)-Abl (treatment of chronic myelogenous leukemia)[118] and gefitinib **2.4** for the inhibition of the epidermal growth factor receptor (treatment of metastatic non-small cell lung cancer)[119] (Figure 2.1). Recently, a second generation of kinase inhibitors that are able to overcome resistance of the first approved drugs have appeared,[120] and new studies have been directed in the development of multi-targeted kinase inhibitors.[121]

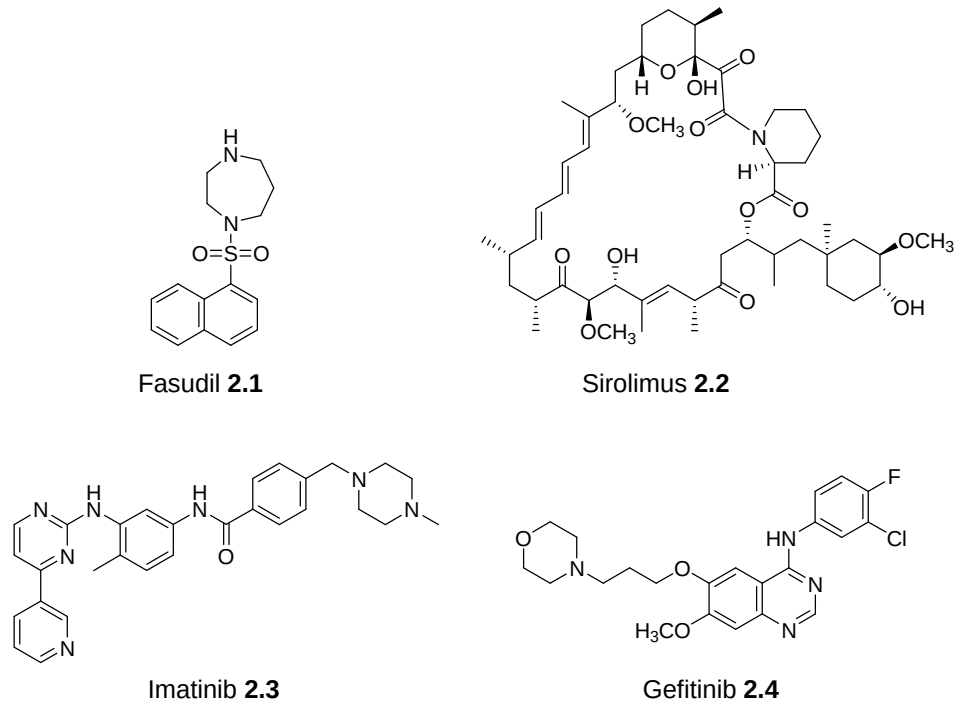


Figure 2.1: Chemical structures of representative small-molecule kinase inhibitors

There is a huge demand for the development of approaches that would help in the design of new kinase inhibitors. A necessity in the area of kinase analysis is the ability to determine and understand the exact substrate specificity of any kinase.

2.3 Protein kinase substrate specificity

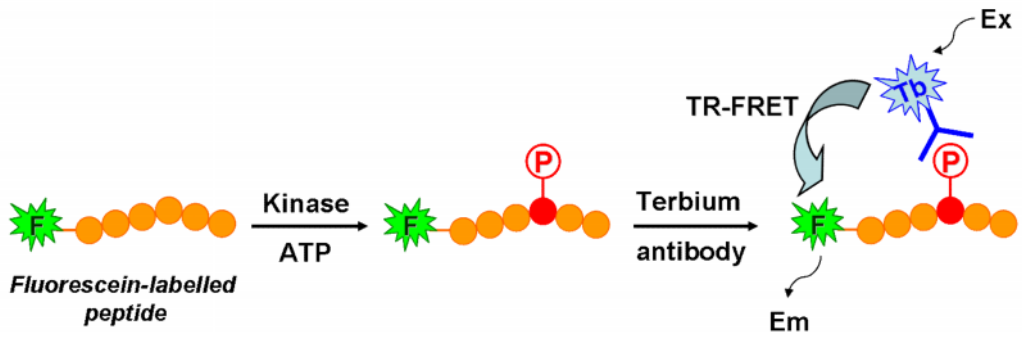
2.3.1 Protein kinase activity detection techniques

A number of well-established technologies that measure kinase activity have been reported.[122-124] They can be classified into three main categories: radioactive, antibody, and non-antibody based methods.

Radioactive-based assays are popular due to their high sensitivity and work with any type of kinase and substrate. In a typical assay, following incubation with a kinase, radiolabelled phosphate (from ^{32}P or ^{33}P ATP) is incorporated into a peptide or protein substrate, and radioactivity corresponding to phosphorylation of the substrate quantified.[122] An alternative that can be applied to a high-throughput screening (HTS) and does not require washing steps is based on a scintillation proximity assay, where a radiolabelled biotinylated substrate is captured by scintillant streptavidin-

coated beads, resulting in the emission of light.[125] However, environmental, safety and cost issues have encouraged the development of non-radioactive approaches such as fluorescence.

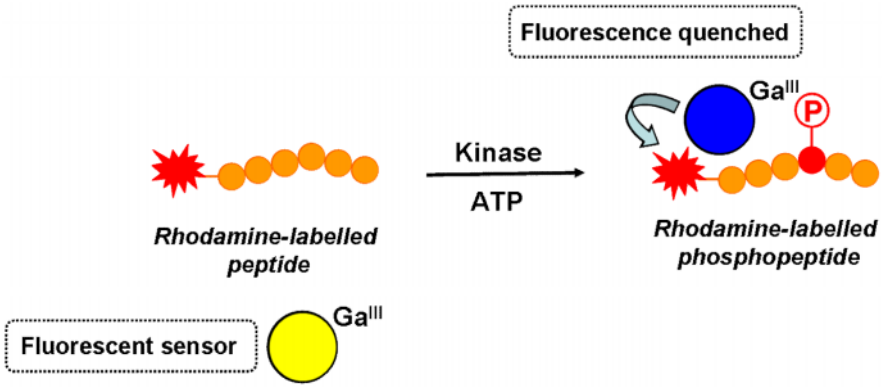
Access to a wide range of commercially available phosphospecific antibodies has favoured the widespread use of antibody/fluorescence-based methods. In a typical experiment, a fluorescently-labelled phosphospecific antibody is used to specifically recognise phosphotyrosine or phosphoserine/phosphothreonine residues.[126] In an enzyme-linked immunosorbent assay, following reaction with a kinase, the phosphorylated peptide is captured on the surface of a 96-well plate, with phosphorylation detected using a phosphospecific primary antibody, or an enzyme-conjugated or fluorescently-labelled secondary antibody.[122] Recently, new methods have been developed such as time-resolved (TR) fluorometry[127] (*e.g.* use of long-lifetime lanthanide labelled antibodies), TR-FRET[128] (*e.g.* use of a terbium-conjugated phosphospecific antibody as a donor and a fluorescein-labelled substrate as an acceptor, Scheme 2.1) and fluorescence polarisation[129] (*e.g.* binding of phosphospecific antibody to a fluorescent phosphorylated peptide, which induces an increase of the compound molecular weight and consequently a reduction in its tumbling time).



Scheme 2.1: Principle of TR-FRET

Recently, alternative non-antibody based methods involving the use of luminescence[130] (*e.g.* quantification of unreacted ATP using a luciferase-based ATP assay) or fluorescence quenching-based assays[131] have started to emerge. In the latter example, following phosphorylation, the phosphate group of a rhodamine-labelled

phosphopeptide interacts with Ga^{3+} metal-ions present on the surface of microspheres coated with a fluorescent polymer (sensor). Thus, polymer and quencher are brought in close proximity resulting in quenching of the polymer fluorescence that was proportional to the level of phosphorylation (Scheme 2.2).



Scheme 2.2: Principle of fluorescence-quenching assays (adapted from Rininsland[131])

2.3.2 Protein kinase profiling methodologies

Over the past decade several kinase profiling methods have been developed and a number of examples are explained in detail below.

In 1995 a solution peptide library (prepared by split and mix methods) was used to determine the substrate specificity of nine protein tyrosine kinases.[132] The general structure of the library used was Met-Ala-X-X-X-X-Tyr-X-X-X-X-Ala-Lys-Lys-Lys, where X represents any amino acid excluding tryptophan, cysteine, tyrosine, serine and threonine. After incubation with a kinase and γ-[^{32}P]-ATP, radiolabelled phosphorylated peptide mixtures (around 1%) were trapped onto a ferric chelating column followed by subsequent Edman sequencing.[133] The abundance of amino acids in the phosphorylated fraction at each position around the phosphorylation site was compared with the abundance of the starting mixture, giving the preferred amino acids at each position. A consensus peptide sequence of optimal substrates was predicted for each kinase (*e.g.* EAIYAAPFAKKK, AEEEIYGEFEAKKK and AEEEEYFELVAKKKK for the protein tyrosine kinases Abl, Src and epidermal growth factor receptor (EGFR) respectively), showing that protein tyrosine kinases have different substrate specificities.

A positional scanning peptide library tagged with biotin at the *C*-terminus, containing nine positions surrounding a central serine/threonine residue was used to profile eight serine/threonine protein kinases.[134] For each sub-library, the fixed positions corresponded to one of the 20 natural amino acids and the variable positions to approximately equimolar mixtures of 17 amino acids excluding cysteine, serine and threonine (Figure 2.2). Additional peptides containing a phosphorylated serine or threonine were also synthesised. The peptide mixtures were treated in parallel with γ-[^{32}P]-ATP and a specific kinase before being immobilised onto avidin-coated membranes allowing quantification by phosphorimaging. The generality of the method was confirmed by the identification of known amino acid sequences surrounding the phosphorylation site using three different classes of serine/threonine kinases (basophilic, acidophilic, and proline-directed kinases). Using this solution-phase approach, results were found to be more accurate than the ones obtained when substrates were first bound to the avidin-coated membrane followed by kinase reaction. Other advantages were the minimal consumption of peptides (100 pmol) and the rapidity of the kinase assay that could be applied to a HTS.

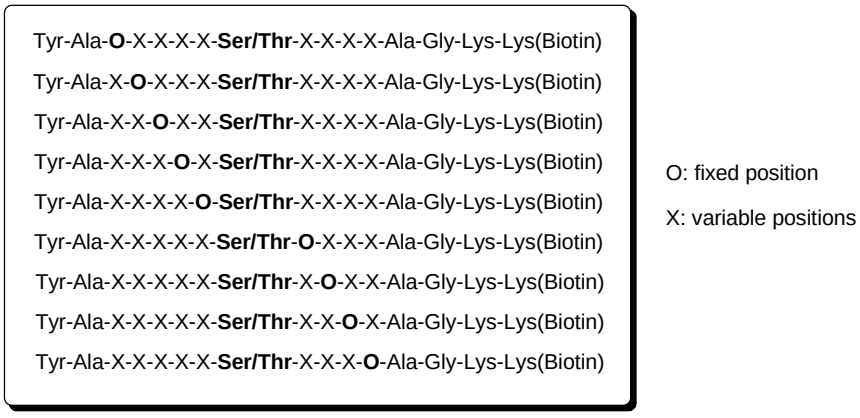


Figure 2.2: Schematic representation of the positional scanning peptide library used by Hutti[134]

Another example was the use of tagged mRNA-peptide and cellular proteomic fusion libraries, where proteins are covalently linked to their encoding mRNA *via* a DNA-puromycin linker (Figure 2.3), to interrogate the non-receptor tyrosine kinase v-Abl.[135] The mRNA peptide fusion library was designed to have a tyrosine residue

surrounded by five random amino acids, and following modification of the library by v-Abl, phosphorylated residues were isolated *via* rounds of immunoprecipitation with the phosphotyrosine-specific α4G10 antibody, and cDNAs were synthesised and amplified by PCR. Sequencing of the cDNAs isolated from the mRNA-peptide fusion library showed that consensus sequences of the substrates were very similar to the known v-Abl motif Ile/Val/Leu-Tyr-X-X-Pro/Phe, where X represented any amino acid.[132] PCR products of selected pools of the mRNA-cellular proteomic fusion library were radiolabelled with [α-^{33}P]-deoxycytidine triphosphate followed by hybridisation onto a cDNA microarray filter, allowing identification of known *in vivo* v-Abl substrates as well as new targets. The tyrosine residues of a few substrates identified in this study were also mutated to phenylalanine, resulting in inhibition of v-Abl kinase activity *in vivo*.

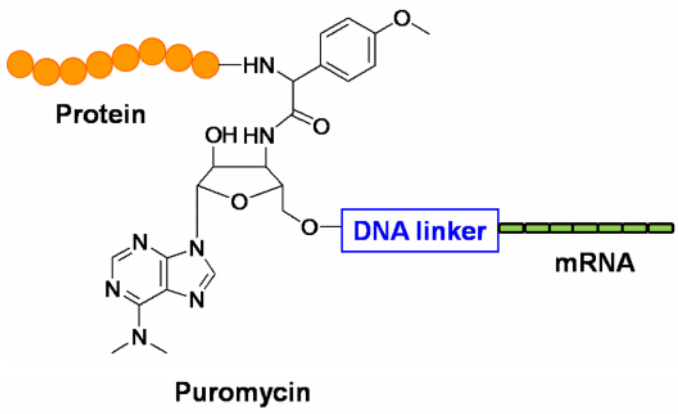
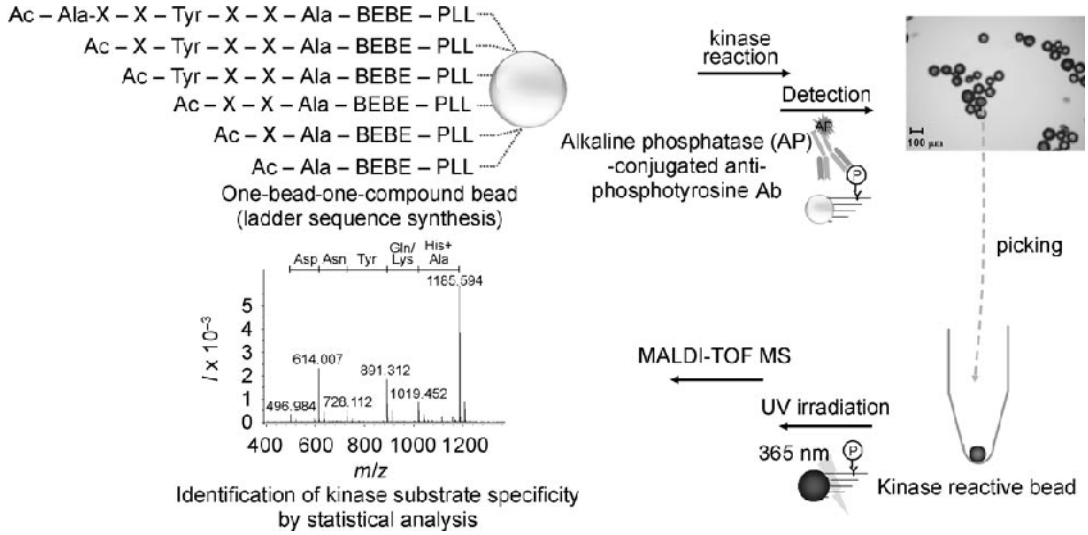


Figure 2.3: Structure of an mRNA-protein fusion

Direct bead-based assays have also allowed the identification of the substrate specificity of two protein tyrosine kinases, p60$^{c\text{-}src}$ and zeta-chain-associated protein kinase 70 using a "one-bead-one-compound"[136] combinatorial peptide library (104,976 compounds).[137] The library was synthesised by ladder synthesis[72] using a split and mix methodology, the general structure of the library contained one fixed tyrosine residue and four randomised positions of 18 amino acids (excluding cysteine and tyrosine) as well as two fixed alanines, a β-alanine-ε-aminocaproic acid-β-alanine-ε-aminocaproic acid (BEBE) spacer and a photolabile linker (PLL). Following kinase reaction, phosphorylated peptides (on beads) were identified using

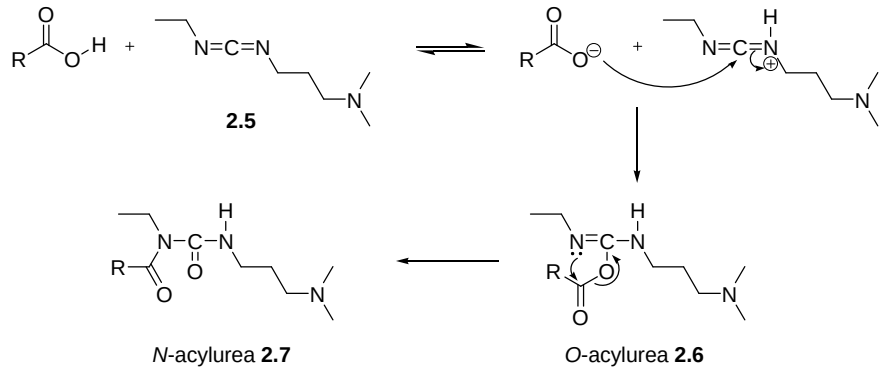
an alkaline phosphatase (AP)-conjugated anti-phosphotyrosine antibody (phosphorylated beads became dark-red following incubation with the AP substrate solution and could be isolated). Photocleavage released peptides from the resin, with peptide sequences identified using ladder based analysis by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) (Scheme 2.3). The same approach was recently applied to screen the Brk protein tyrosine kinase.[138]



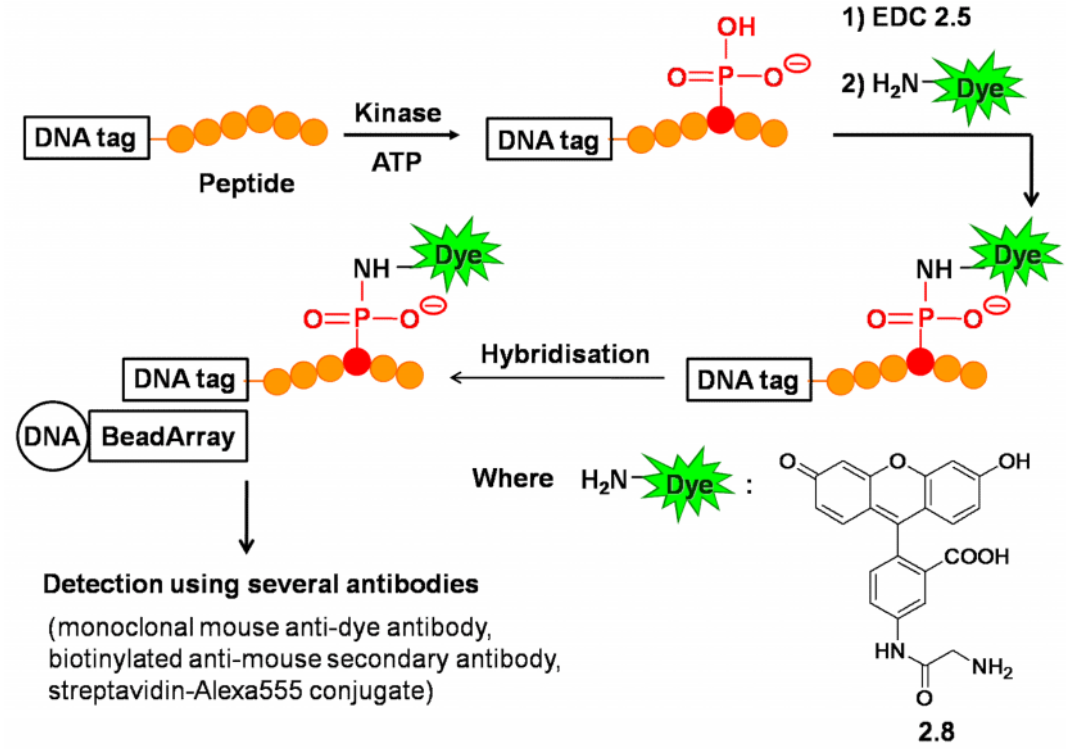
Scheme 2.3: HT identification of kinase substrate specificity using a one-bead-one-compound library (reproduced with permission)[137]

Two combinatorial pentapeptide and heptapeptide libraries on beads (one peptide per bead) were used to screen for substrates of the serine/threonine protein kinase PKA. Following incubation with PKA and γ-[^{32}P]-ATP, radiolabelled beads were isolated and sequenced to identify the well-known "RRXS" peptide motif for PKA (where X was any amino acid).[139] Another bead-based platform has been used to study phosphorylation of 26 different kinases.[140] Assays were carried out using 900 different peptide substrates that were synthesised in parallel and encoded by a unique oligonucleotide tag. Following incubation with a kinase or a mixture of kinases, phosphorylated peptides were detected through chemical phosphate activation and amine-functionalised dye coupling using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC) **2.5**. EDC catalyses the coupling of phosphates but also carboxylic acids to primary amines. To selectively label the phosphorylated

residue, the authors had to first block the carboxylic groups of the amino acid side chains using EDC at low pH, which induced formation of unstable *O*-acylurea intermediates **2.6** that rearranged to the stable (unreactive) *N*-acylurea products **2.7** (Scheme 2.4). DNA-encoded phosphorylated peptides could then be selectively labelled using fluorescein glycine amide **2.8**, and hybridised to complementary DNA sequences attached to individual beads on a BeadArray (Illumina), allowing identification of the substrates recognised by the kinase. Surprisingly, dyes were then detected using several antibodies, presumably due to the low level of labelling (Scheme 2.5). The main advantages of this approach relied on the fact that tyrosine-, serine- and threonine-phosphorylated peptides could be successfully detected without the use of radiolabelled-ATP or phosphospecific antibodies.

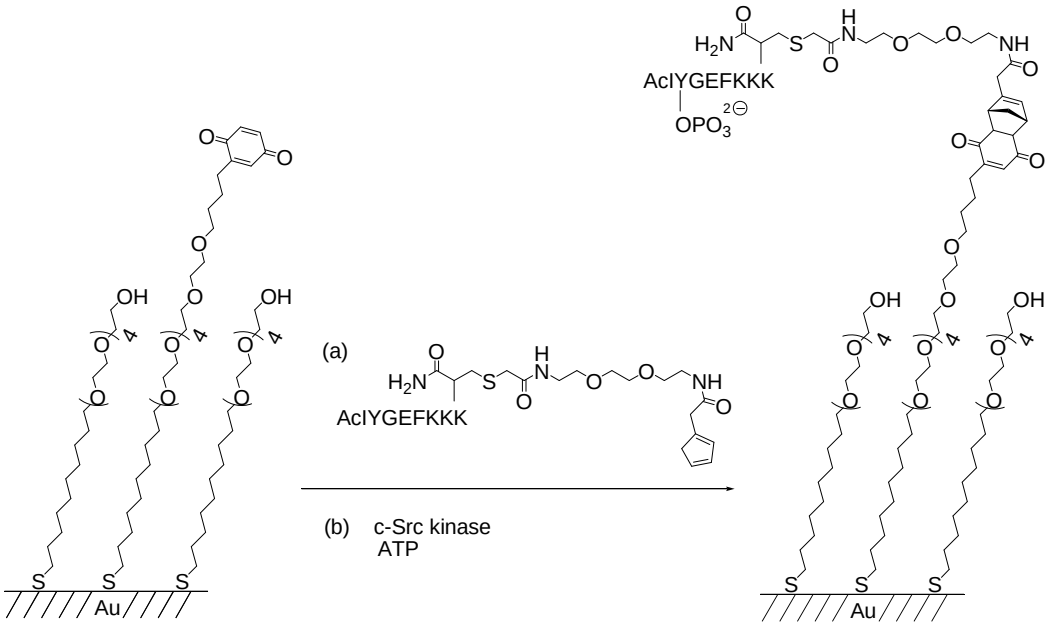


Scheme 2.4: Coupling using EDC



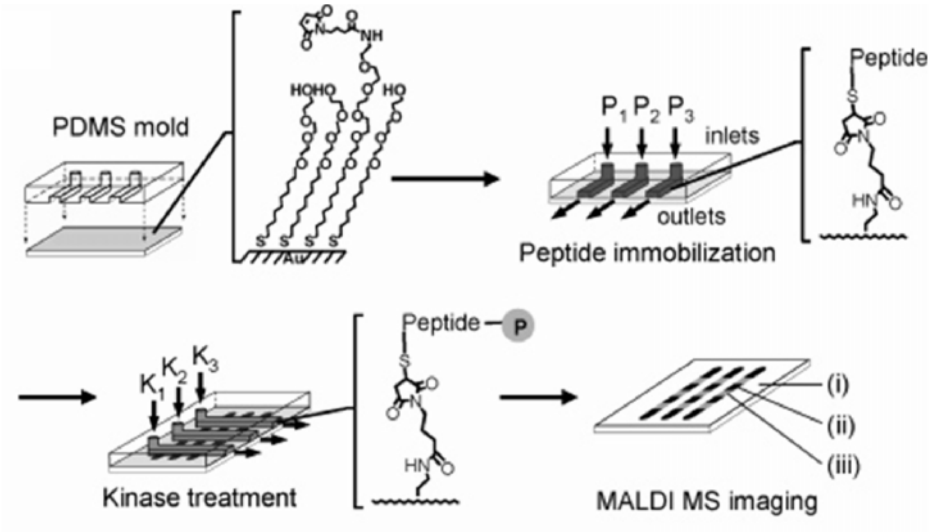
Scheme 2.5: Protein kinase assay developed by Shults[140]

Several microarray-based methods have also been developed.[141] For example, kinase substrates (cyclopentadiene-capped peptides) were immobilised *via* a Diels-Alder reaction on self-assembled monolayers (SAMs) containing benzoquinone groups on the surface. Inert surfaces of SAMs prevented unspecific binding with proteins or radioisotopes, thus eliminating the requirement of blocking steps. Immobilised peptides were uniformly accessible to kinases, with phosphorylation detected using surface plasmon resonance, fluorescence and phosphorimaging. This approach was used to determine the non-receptor tyrosine kinase c-Src activity and evaluate three Src inhibitors (Scheme 2.6).[86]



Scheme 2.6: Immobilisation of the optimal c-Src kinase substrate on a SAM followed by on-chip phosphorylation

Similarly peptides containing a *C*-terminal cysteine were immobilised on SAMs displaying maleimide groups on gold surfaces to study kinases and quantitatively evaluate inhibition of these kinases by known inhibitors. Kinase assays simply required addition of a specific kinase on SAMs followed by treatment with the 2,4,6-trihydroxyacetophenone matrix, and analysis by MALDI-TOF MS, although only seven peptides were studied.[142] The same group also adapted this approach to assay multiple enzyme (seven kinases and four phosphatases) activities by combining microfluidic networks and peptide arrays. In that case, a poly(dimethylsiloxane) (PDMS) stamp with six microchannels was sealed on top of the SAM, solutions of peptide substrates were filled into each channel and added on the SAM. Following washings, the PDMS stamp was reapplied to the SAM in a perpendicular way so that each channel crossed the six parallel lines of immobilised peptides. Finally, solutions containing different enzymes were injected into each separate channel and incubated on the SAM, before subsequent MALDI-TOF MS analysis (Scheme 2.7).[143]



Scheme 2.7: Use of microfluidic networks and peptide arrays to assay multiple kinases (reproduced with permission)[143]

Recently, another label-free method that uses cysteine-terminated peptide-conjugated gold nanoparticles attached to an amine-functionalised glass surface and matrix-free secondary-ion mass spectrometry analysis has been reported to determine Abl and PKA specificity.[144] Microarrays of 710 or 1394 peptides (13-mers), generated by SPOT synthesis,[55] with subsequent oxime attachment to a glass slide and phosphorylation by a kinase in the presence of radioactive ATP allowed "phosphorimaging" and analysis of PKA (Figure 2.4) and 3-phosphoinositide-dependent kinase specificity respectively.[102] The same strategy was applied to determine the preferred substrates of casein kinase 2, and for Abl protein tyrosine kinase, preferred substrates were identified using a fluorescein-labelled anti-phosphotyrosine antibody, giving results comparable to those obtained using radioactivity.[145] It was also shown that Abl phosphorylation sites could be predicted using a peptide microarray containing 1433 peptides.[146]

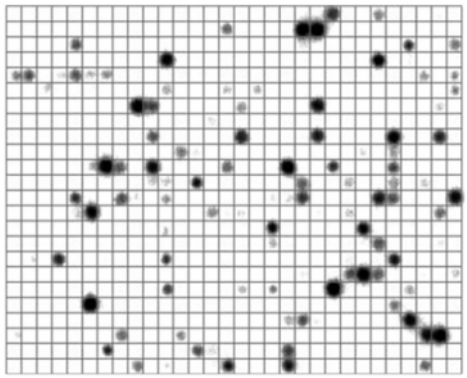


Figure 2.4: A peptide microarray image obtained after incubation with PKA and ^{32}P-ATP (reproduced with permission)[102]

A 125-member PNA-encoded library (Figure 2.5) has been incubated with the protein tyrosine kinase Abl, before being immobilised on a DNA microarray and treated with a fluorescein isothiocyanate-labelled anti-phosphotyrosine antibody. Encoding of each randomised amino acid by a specific PNA triplet enabled rapid and easy identification of peptide sequences phosphorylated by the kinase Abl.[81] Later, Harris reported the use of PNA-encoded peptides to study kinase activity in a similar way but results were acquired using a very small number of peptides (nine substrates).[77]

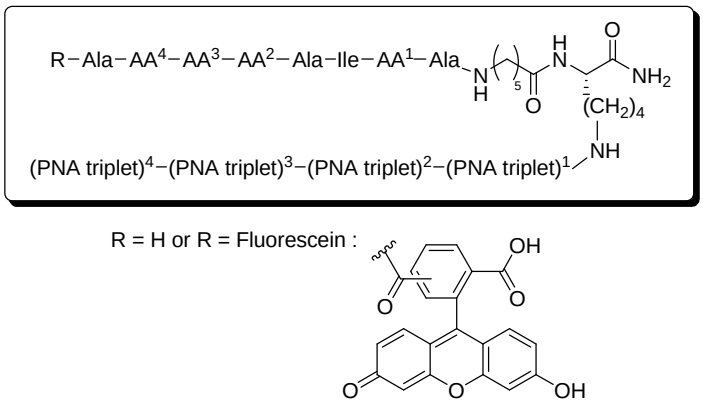


Figure 2.5: General structure of the 125-member pool of a PNA peptide library

2.3.3 Aim of the project

The huge potential of using PNA-encoded libraries to profile a specific protein kinase has been proven.[77, 81] However, several factors such as different PNA melting

temperatures, variations in concentration and hybridisation efficiency of the different library members had to be taken into consideration. To have robust and reliable data, it became essential to design PNA-encoded libraries that contained internal controls. This was successfully achieved by Bradley who introduced the dual colour concept.[82] This relied on a FRET system, composed of the donor fluorescein and the quencher TAMRA, integrated within the PNA-encoded library. This technique was efficiently applied to interrogate 10,000 protease substrates. The aim of the project was to transpose this dual colour approach to determine protein kinase substrate specificity *via* the use of a specifically designed 10,000-member split-and-mix PNA-encoded FRET based peptide library, allowing quantification of phosphorylation.

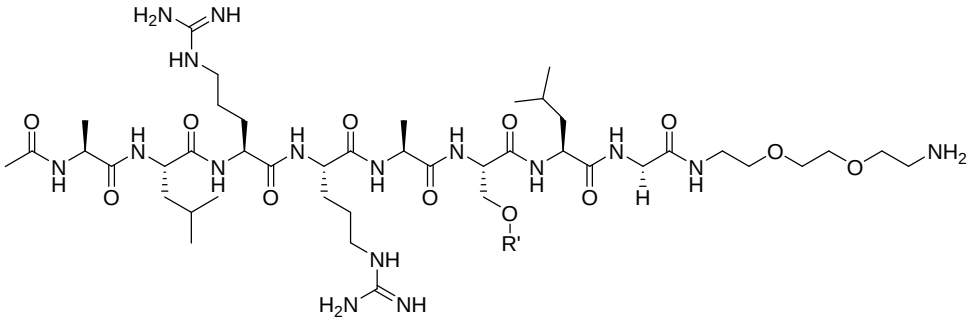
2.4 Optimisation of the detection methodology

2.4.1 Introduction

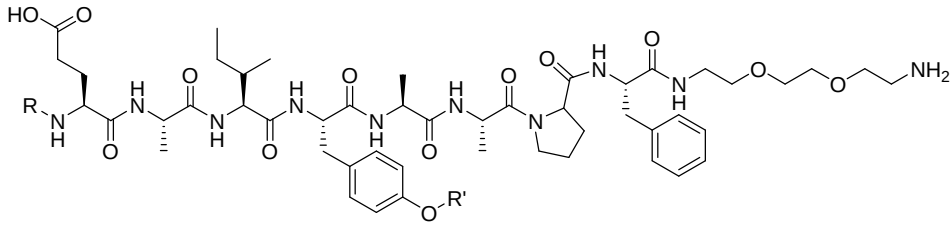
Despite the large number of available methods capable of detecting phosphorylated peptides, it remains an important and difficult task to select the correct one. To develop a general and valuable approach, the use of a non-radioactive antibody-based methodology that could be applied to HTS appeared an attractive strategy. Yao used fluorescently-labelled antibodies for the detection of phosphorylated tyrosine and serine.[126] Moreover, the detection of phosphorylated peptides within a PNA-encoded library using a fluorescein isothiocyanate-labelled anti-phosphotyrosine antibody has proven efficient.[81] However, one remaining issue implied by the use of fluorescently labelled antibodies is the possible lack of specificity that can produce false positives and interfere in the interpretation of the results. Therefore, it was necessary to prove that the selected fluorescently labelled antibodies specifically recognised tyrosine- or serine-phosphorylated peptides and not any particular amino acid or peptide motif. Although the final objective was to apply this detection method to peptide-PNA conjugates, control experiments were first carried out using peptide microarrays, containing some tyrosine/serine (non)phosphorylated peptides. This allowed antibody binding selectivity as well as optimisation of antibody titres. Regarding possible interactions of antibodies with PNA tags, an additional experiment was carried out with an unmodified PNA-encoded library (no incubation with a kinase and ATP) and is reported in section 2.5.2.

2.4.2 Syntheses of control and phosphorylated peptides

To prove that fluorescently labelled antibodies were specific to tyrosine or serine residues, known kinase substrates of the protein tyrosine kinase Abl (EAIYAAPF)[132] and protein serine kinase PKA (ALRRASLG known as Kemptide)[147] as well as their corresponding phosphorylated peptides were synthesised. Four acetyl-capped peptides [Ac-ALRRASLG-PEG-NH_2 **2.9**, Ac-ALRRApSLG-PEG-NH_2 **2.10**, Ac-EAIYAAPF-PEG-NH_2 **2.11** and Ac-EAIpYAAPF-PEG-NH_2 **2.12**, where PEG = $(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_2$] were thus assembled *via* solid phase synthesis using a 2-chlorotrityl chloride resin and Fmoc-protected amino acids. An additional peptide labelled with 5(6)-carboxyfluorescein (Fluorescein-EAIYAAPF-PEG-NH_2 **2.13**) was synthesised as a control peptide (Figure 2.6).



R' = H: **2.9** or R' = PO_3H_2: **2.10**

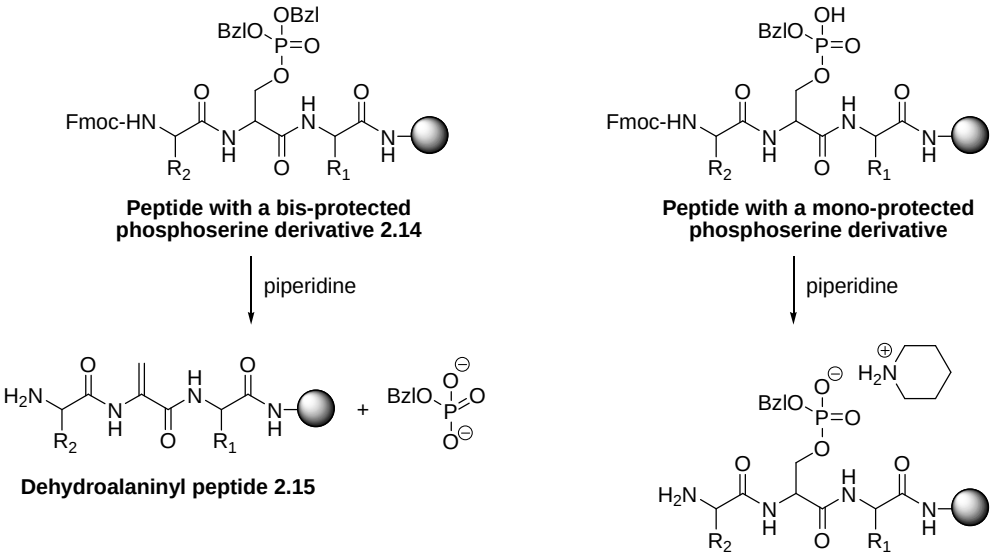


R = Ac, R' = H: **2.11** or R' = PO_3H_2: **2.12**
R = Fluorescein, R' = H: **2.13**

Figure 2.6: Chemical structures of the control and phosphorylated peptides

To synthesise phosphorylated peptides, preformed protected phosphoamino acids (Fmoc-Tyr(PO(OBzl)OH)-OH and Fmoc-Ser(PO(OBzl)OH)-OH) were used. During Fmoc deprotection, the addition of piperidine onto a peptide containing a phosphoserine with a bis-protected phosphate **2.14** can lead to rapid β-elimination of

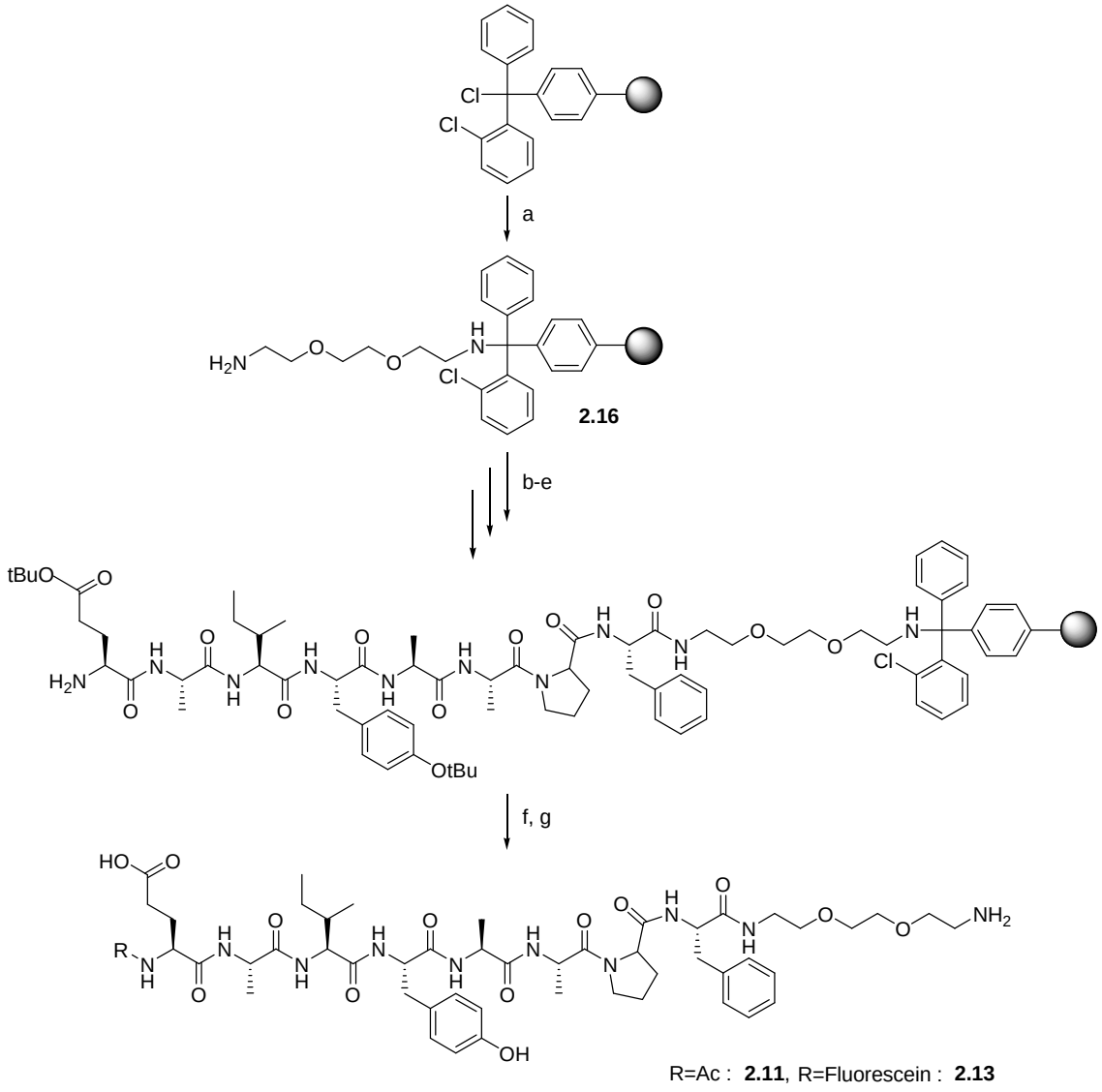
the phosphate group and formation of the corresponding dehydroalaninyl peptide **2.15**.[148] Wakamiya showed that this side reaction can be almost totally eliminated using a mono-protected phosphoserine derivative (*e.g.* Fmoc-Ser(PO(OBzl)(OH)-OH) due to ionisation of the phosphate group that inhibits the β-elimination (Scheme 2.8).[149] Perich tested different uronium, BOP or DIC-based coupling methods and reported HBTU/HOBt/DIPEA, HATU/HOAt/DIPEA as the most efficient coupling procedures to incorporate these building blocks. A classical coupling reagent such as DIC resulted in a higher proportion of deleted peptides due to a possible participation of the phosphate hydroxyl group during the activation step.[150] Thus, phosphorylated and non-phosphorylated peptides were synthesised using HBTU/HOBt/DIPEA.



Scheme 2.8: Stability of phosphoserine derivatives to piperidine

Since the peptides needed to be immobilised onto glass slides, a spacer (2-[2-(2-aminoethoxy)ethoxy]ethanamine) was first attached onto the resin to give **2.16**. Syntheses of the known Abl substrate **2.11** and the fluorescein-labelled peptide **2.13** were carried out using a CEM Liberty microwave peptide synthesiser that allowed increased coupling efficiency, reduced reaction times and automation of the washing steps. Fmoc deprotections were carried out at 25 °C and peptide couplings were performed under microwave irradiation at 60 °C for 20 min (double coupling). Peptides were capped with either an acetyl group to give only one free amine for

immobilisation or 5(6)-carboxyfluorescein (fluorescein) for labelling. The formation of various side products during the in-situ activation of fluorescein has been reported by Fischer,[151] but these impurities could be easily eliminated by treating the resins with 20% piperidine in DMF prior to cleavage. Finally, peptides were released from the resin using a solution of TFA/TIS/DCM (50/45/5) for 2 h and purified by preparative HPLC (Scheme 2.9).

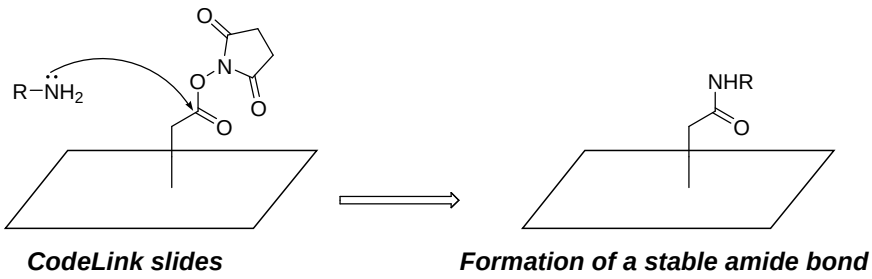


Scheme 2.9: Microwave-assisted preparations of Ac-EAIYAAPF-NH-PEG-NH_2 **2.11** and Fluorescein-EAIYAAPF-NH-PEG-NH_2 **2.13**. Reagents and conditions: (a) 2-[2-(2-aminoethoxy)ethoxy]ethanamine (10 eq), dry THF, 16 h; (b) 20% piperidine in DMF, 2 cycles of 7 min, 25 °C; (c) Fmoc-AA-OH (5 eq), HBTU/HOBt (5 eq), DMF (0.2 M), 20% DIPEA in NMP (10 eq), 20 min, 60 °C; (d) Repeat (c); (e) Repeat (b)-(d); (f) Ac_2O/pyridine (50/50), 1.5 h when R=Ac or 5(6)-carboxyfluorescein (5 eq), HBTU/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq) followed by washings with 20% piperidine in DMF, 2 cycles of 10 min when R=Fluorescein; (g) TFA/DCM/TIS (50/45/5), 2 h

However, since the minimum working scale for the microwave peptide synthesiser was 100 mg of resin, the phosphorylated peptides were synthesised manually at 25 °C using 25 mg of resin. Regarding Kemptide **2.9** synthesis, some deleted peptides (one missing arginine) could be observed using the microwave peptide synthesiser; the advantage of manual synthesis was that coupling of arginine could be repeated several times until completion by running a ninhydrin test (see section 4.1.2). The use of Wang linker instead of 2-chlorotrityl chloride resin didn't improve the purity of the crude compound. The main problems for this peptide were the partial deprotection of the 2,2,4,6,7-pentamethyldihydro-benzofuran-5-sulfonyl (Pbf) group on arginine. Due to the presence of two consecutive arginines, the reagent K cocktail (TFA/H_2O/phenol/thioanisole/EDT: 82.5/5/5/5/2.5) had to be used to fully remove the two Pbf groups.

2.4.3 Printing onto CodeLink slides

CodeLink slides (Amersham Biosciences) are coated with a hydrophilic polymer containing *N*-hydroxysuccinimidyl (NHS) esters that can covalently immobilise amine-containing compounds (Scheme 2.10).



Scheme 2.10: Immobilisation of amine-modified compounds onto CodeLink slides

Solutions containing the peptides Ac-EAIYAAPF-PEG-NH_2 **2.11**, Ac-ALRRASLG-PEG-NH_2 **2.9**, Ac-EAIpYAAPF-PEG-NH_2 **2.12**, Ac-ALRRApSLG-PEG-NH_2 **2.10** and Fluorescein-EAIYAAPF-PEG-NH_2 **2.13** at 8 different final concentrations (A=5 mM, B=1 mM, C=0.5 mM, D=0.1 mM, E=50 µM, F=10 µM, G=1 µM and H=100 nM) in printing buffer were printed onto CodeLink slides in region 1, 2, 3, 4 and 5 respectively (16 replicates, 4x4 sub-arrays) (Figure 2.7). Using fluorescein filters,

peptide and phosphorylated peptide printing was confirmed since fluorescent spots for the control peptide **2.13** (region 5) could be observed (Figure 2.7).

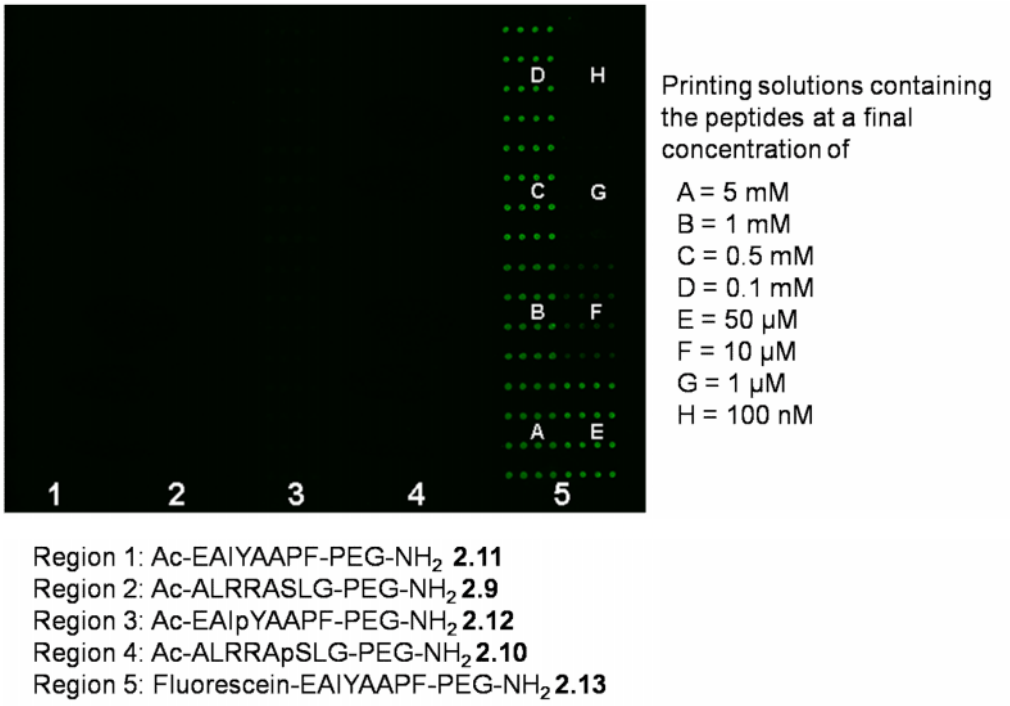


Figure 2.7: Image of the peptide microarray obtained using a fluorescein filter (no background reprocessing)

2.4.4 Detection of tyrosine-phosphorylated peptides

Cyanine dyes have a wide range of biological applications including labelling of nucleic acids or proteins.[152] Sulfonated Cy3 is a water soluble highly fluorescent dye belonging to the cyanine dye family that can be detected using a rhodamine filter. FluoroLink MAb Cy3 labelling kit (Amersham Biosciences) was used to label a commercially available monoclonal anti-phosphotyrosine (mouse IgG2b isotype) clone PY-20 antibody (Sigma). The Cy3 dye was synthesised with an NHS ester that could react with the protein *via* free amino groups (Figure 2.8).

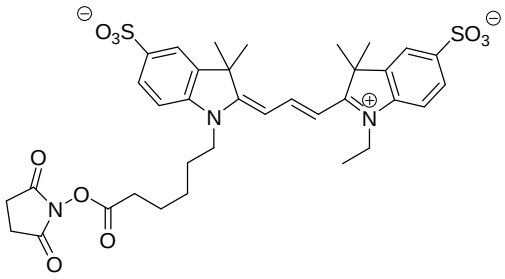
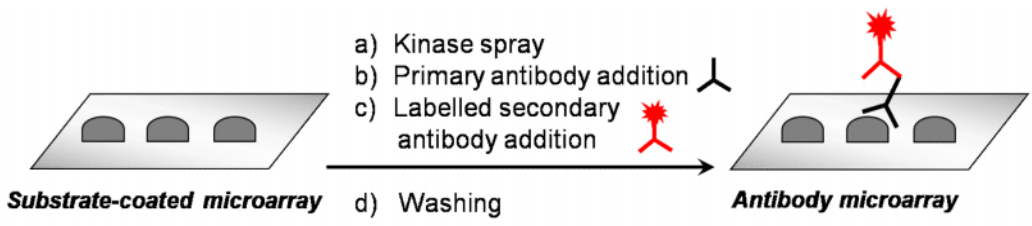


Figure 2.8: Sulfonated activated Cy3

Cy3 was conjugated to the antibody, and unconjugated dye was separated from the labelled antibody by gel filtration. The ratio of dye/protein corresponding to the average number of dye molecules coupled to each protein molecule was calculated by determining the dye concentration using the extinction coefficient of the Cy3 dye at 552 nm and the absorbance of labelled protein at 280 nm. The calculated ratio of dye/protein was really low (values are usually between 4 and 12, but a value of 1.2 was obtained), indicating a low labelling of the antibody. Two factors such as protein concentration and reaction pH may have influenced the labelling efficiency of the reaction.

The method described by Diamond was chosen as an alternative strategy.[153] In this approach, protein active slides were first coated with a specific kinase substrate. Small molecules solubilised in 10% glycerol kinase reaction buffer along with ATP were printed on substrate-coated slides followed by activation with a kinase that was aerosol sprayed. Finally, slides were treated with a primary antibody and an Alexa Fluor 555-labelled secondary antibody, allowing amplification of the signal. This detection method was applied to the screening of more than 20 serine/threonine and tyrosine kinases (Scheme 2.11).



Scheme 2.11: Detection method using primary and labelled secondary antibodies

Thus, for the peptide microarray experiment, a monoclonal anti-phosphotyrosine (mouse IgG2b isotype) clone PY-20 antibody (Sigma) was used as a primary antibody and a Cy3-goat anti-mouse IgG (H+L) conjugate (Invitrogen) as a secondary antibody. Before antibody addition on the peptide microarray, it was essential to block the chip using 0.05% Tween-20 and phosphate buffered saline (TPBS)/ 0.1% Bovine Serum Albumin (BSA) to avoid non-specific binding. The ratios of primary and secondary antibodies were then optimised. After blocking the chip, addition of an anti-phosphotyrosine antibody (diluted 1/500) and a Cy3-labelled secondary antibody (diluted 1/1000), the chip was scanned using a Cy3 filter (Figure 2.9).

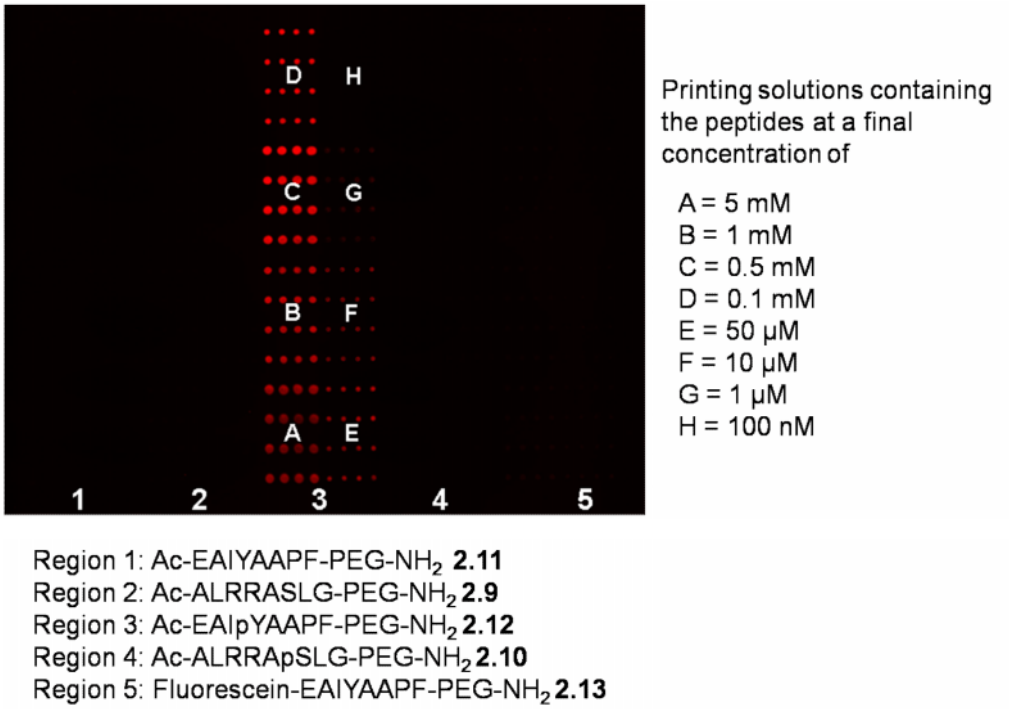


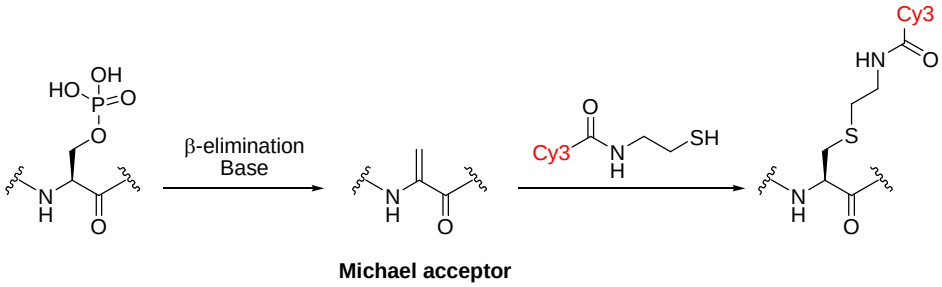
Figure 2.9: Image of the peptide microarray obtained using a Cy3 filter (no background reprocessing)

Only the tyrosine-phosphorylated peptide **2.12** immobilised on the slide (printed in region 3) could be detected, confirming the high specificity of the pair of antibodies against the phosphorylated tyrosine residue. A decrease in fluorescence intensity was observed as the concentration of the printing solution containing **2.12** reduced from 0.5 mM to 100 nM. However, for higher printing solution concentrations (1 mM and

5 mM), bigger spots were noticed and the fluorescence intensity appeared to decrease due to self-quenching of the Cy3 dye. This method was really sensitive since a 1 μM final concentration of the printing solution containing **2.12** could be readily detected.

2.4.5 Detection of serine-phosphorylated peptides

The same procedure was applied and optimised to detect serine-phosphorylated peptides but using a monoclonal anti-phosphoserine (mouse IgG1 isotype) clone PSR-45 antibody (Sigma) as a primary antibody and the same Cy3-labelled secondary antibody. The pair of antibodies bound to all immobilised peptides so that it was not possible to specifically detect serine-phosphorylated peptides. False-negative results and high cross-reactivity were the main drawbacks associated with the use of the anti-phosphoserine/threonine antibodies.[145] Unlike the anti-phosphotyrosine antibodies that recognised a phosphorylated tyrosine independently of adjacent amino acids, anti-phosphoserine/threonine antibodies were sequence dependent. Thus, the development of a custom antibody for a specific protein serine kinase is often required. An alternative could be to use a small molecule fluorophore phosphosensor such as the Pro-Q Diamond dye (λ_{ex} = 555 nm and λ_{em} = 580 nm) that interacts selectively with phosphorylated residues and does not rely upon any sequence context. Tyrosine-, serine- and threonine-phosphorylated peptides immobilised on microarrays were efficiently detected following incubation of the slides in Pro-Q Diamond phosphoprotein stain.[154] Other strategies involve β-elimination under basic conditions of a serine-phosphorylated peptide followed by addition of a fluorescently labelled thiol[155] (Scheme 2.12) or adaptations of the method described by Shults (see section 2.3.2).[140] In the latter case, following incubation with a protein serine kinase and ATP, the phosphate group of PNA-encoded serine-phosphorylated peptides would be activated using EDC **2.5** with subsequent selective labelling using an amine-functionalised dye. Following hybridisation of these fluorescently-labelled PNA-encoded serine-phosphorylated peptides on a DNA microarray, the substrates recognised by the kinase would be readily identified.



Scheme 2.12: Nucleophilic attack by Cy3-labelled thiol on dehydroserine

Several attempts to carry out on-chip β-eliminations of a serine-phosphorylated peptide using $Ba(OH)_2$[156] or NaOH[155] as bases followed by addition of a fluorescein-labelled thiol synthesised in the laboratory were unsuccessful (peptide microarrays printed in section 2.4.3 were used). Due to the silane base coating of the CodeLink slides, it was not recommended to incubate slides in solutions with a pH greater than 9. NaOH solutions (pH > 9) were used to optimise β-eliminations of serine-phosphorylated peptides, this might have resulted in base-mediated hydrolysis of the Si-O-Si bond and damage to the surface of the slide. Moreover a very high background was obtained after addition of the labelled thiol. The same negative results were obtained using aldehyde slides.

In conclusion, it was not possible to detect selectively serine-phosphorylated peptides immobilised on a microarray. However, detection of tyrosine-phosphorylated peptides using a pair of antibodies (anti-phosphotyrosine and Cy3-labelled secondary antibodies) was successfully achieved. Future experiments to determine kinase substrate specificity will thus only focus on the profiling of protein tyrosine kinases.

2.5 Determination of the kinase substrate specificity from a 10,000-member PNA-encoded library

2.5.1 General structure of the library

Herein is described an approach that allows the screening and analysis of 10,000 kinase substrates in a single experiment, with analysis of ALL 10,000 members of the library on a one by-one basis *via* the combined application of DNA microarrays, fluorescently labelled secondary antibodies and PNA-split and mix encoding.[78, 81, 82,

[157] This was achieved by the split and mix synthesis of a 10,000-member PNA-encoded library containing the peptide sequence -Phe-Gln-AA^4-AA^3-Tyr-AA^2-AA^1-Ile-Lys- (Figure 2.10), with the expected phosphorylation site surrounded by four variable positions (AA^1, AA^2, AA^3 and AA^4), each containing 10 different amino acids. This library was synthesised by Bialy and Díaz-Mochón.[158]

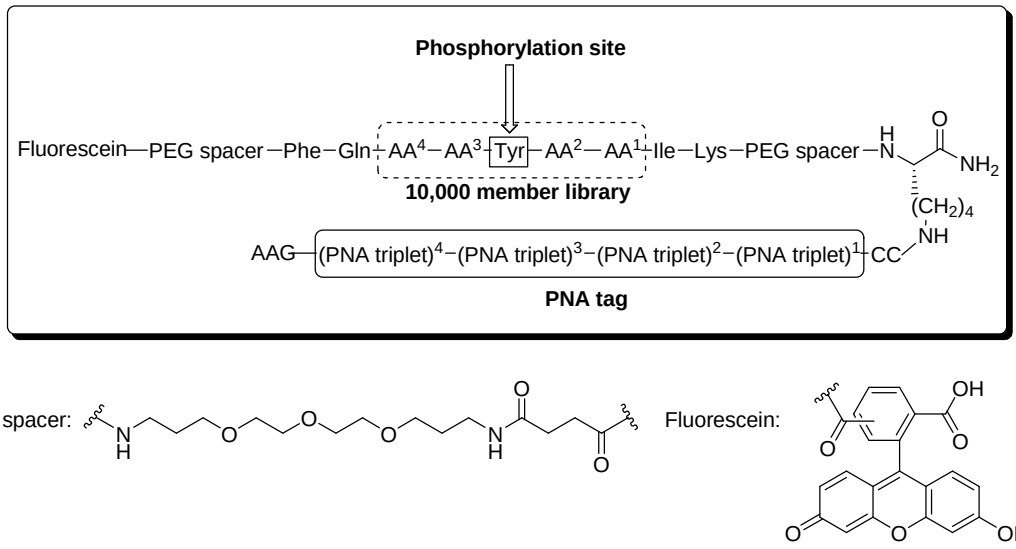


Figure 2.10: General structure of the 10,000-member PNA-encoded library

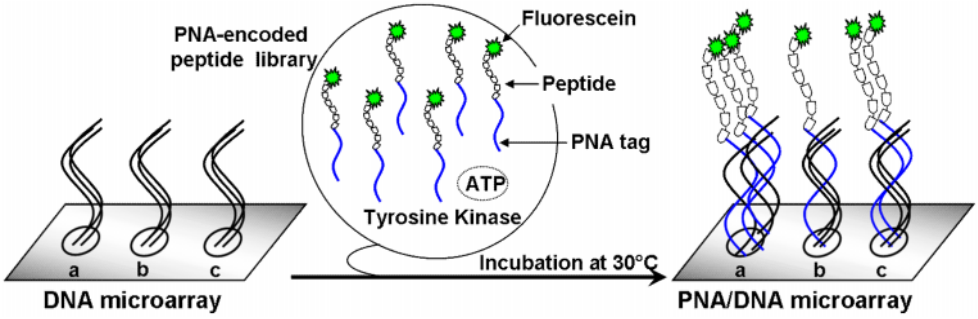
The library contained both natural and unnatural amino acids with differing properties (hydrophobic, hydrophilic, neutral, basic and acidic); Tyr was not included in the randomised positions to avoid multiple phosphorylation sites. Each amino acid was encoded by a specific PNA triplet (Table 2.1) to give an encoding PNA tag composed of a PNA dodecamer. The orthogonality and robustness of the PNA and peptide split and mix chemistry have been reported previously.[4, 20, 36] The PNA tag was also supplemented with two common PNA sequences at the *C*-terminus (CC) and three at the *N*-terminus (AAG) to help control the selectivity of hybridisation and reduce mismatches.[159] Two PEG spacers were also added between the peptide arm and the PNA oligomer to distance the peptide away from the PNA strand and finally, every peptide was labelled with fluorescein in order to allow internal control of each and every library member.[81, 82]

Table 2.1: The PNA tags used to encode amino acids of the library

PNA triplet (from N → C)	AA^4, AA^2, AA^1	AA^3
CTT	Ile	Ile
GTT	Val	Val
CTC	Phe	Ala
ATC	Pro	Pro
GTC	Arg	Arg
GTA	Glu	Glu
GTG	Lys	Lys
ACA	D-Pro	D-Ala
GCA	Ser	Ser
GCG	D-Val	Pro

2.5.2 Kinase assay

Following solid-phase synthesis, the library was incubated with the Abelson non-receptor tyrosine kinase (Abl)[160] and ATP allowing phosphorylation of any peptides recognised by Abl, with subsequent pull down (hybridisation) of the 10,000-member PNA-encoded library onto the DNA microarray. In this approach, all 10,000 peptides are in essence "delivered" to a specific address on the array by virtue of their specific and unique tag (zip or post-code) *via* anti-parallel PNA/DNA duplex formation (Scheme 2.13).[81, 82]



Scheme 2.13: Hybridisation of the library

The arrays used contained 22,575 features with 10,000 oligonucleotide sequences complementary to the PNA sequences (each in duplicate) and 2,575 control DNA

sequences that were designed to be non-complementary to any of the PNA tags in the library. Analysis of the image obtained after scanning the chip using a fluorescein filter showed a total absence of fluorescence on the 2,575 control sequences [*e.g.* empty white circles (see insert)], confirming the highly selective hybridisation (Figure 2.11). Every fluorescent spot represents a unique peptide, whose sequence is known *via* its PNA tag and location on the array. Note the variation in fluorescence intensity which highlights differences in hybridisation and synthesis efficiency that has to be controlled using relative ratios of fluorescence rather than absolute values.

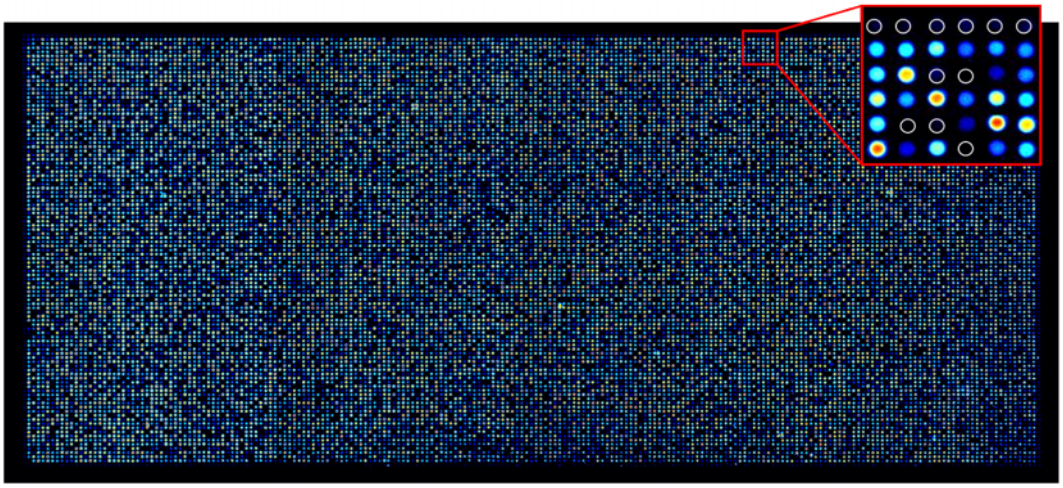


Figure 2.11: Image of the 22,575-member custom DNA microarray using a fluorescein filter

In essence the DNA array has been converted into a peptide array containing some phosphorylated peptides. These were subsequently detected using a primary/secondary antibody approach as explained in section 2.4.4. Therefore, an anti-phosphotyrosine antibody (mouse IgG, diluted 1/500) was added on the array followed by addition of a Cy3-labelled secondary antibody (anti-mouse IgG, diluted 1/500).[153] The dilution factor of the secondary antibody was higher than the one optimised for peptide microarrays in order to simply obtain a stronger Cy3 fluorescence signal after scanning. Any phosphorylated peptides ["hits", *e.g.* red spots in white circles, Figure 2.12 (see insert)], localised onto its defined-DNA, were readily identified by scanning the DNA microarray using a Cy3 filter (Figure 2.12).

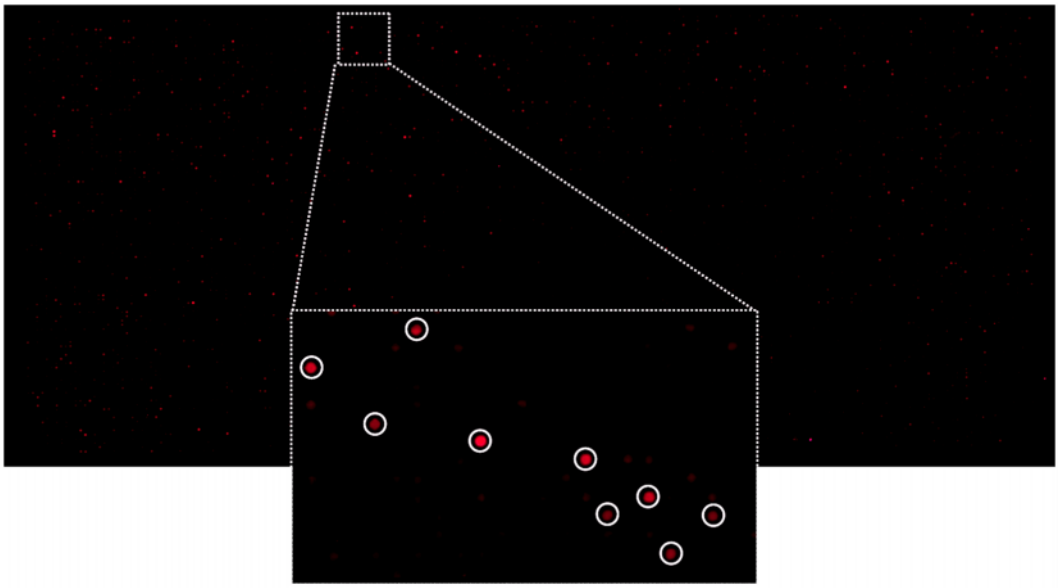


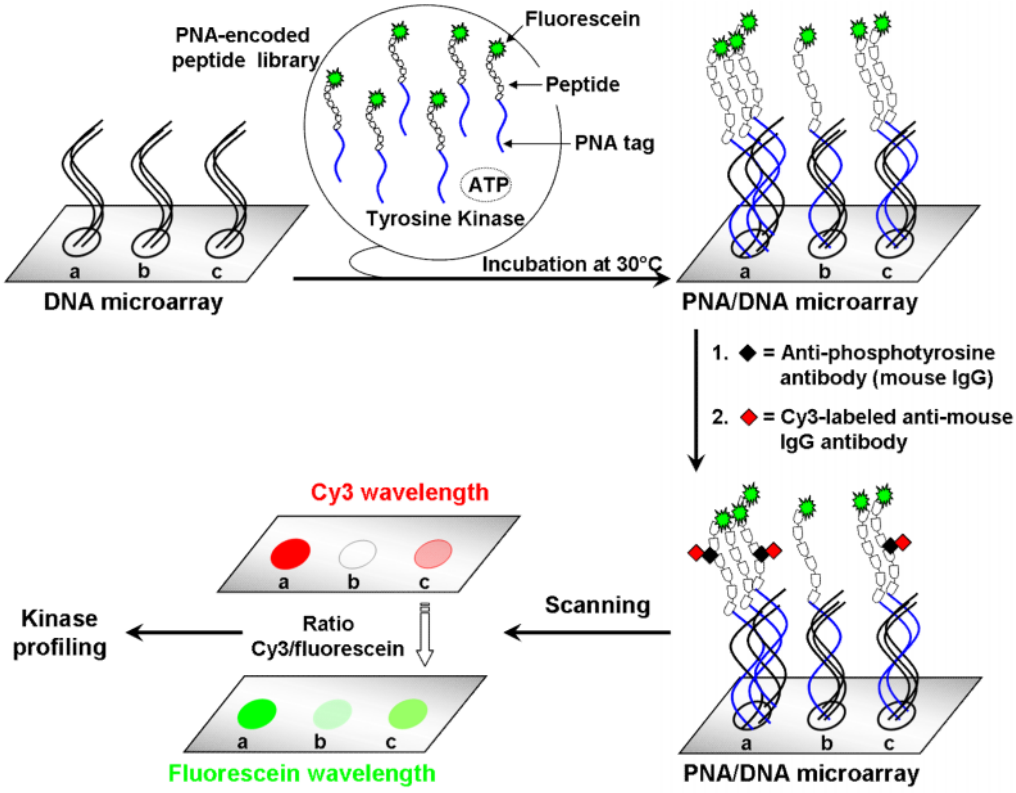
Figure 2.12: Image of the 22,575-member custom DNA microarray using a Cy3 filter.

To verify that the pair of antibodies had no affinity with any PNA tag from the library, an unmodified 10,000-member PNA-encoded library (no incubation with Abl and ATP) was hybridised on the DNA chip. After incubation (in a blocking buffer), the chip was treated with the primary antibody followed by addition of the secondary labelled antibody. No fluorescence was observed, confirming once again the high specificity of this pair of antibodies (data not reported).

2.5.3 Dual control approach

Since every peptide was labelled with fluorescein, the resulting ratio of Cy3/fluorescein provided an internal control for each and every member of the library, which was crucial due to natural variations in melting temperatures, concentrations, as well as differences in hybridisation efficiency between the 10,000 different library members. The use of just Cy3 fluorescence for example would mean it would be impossible to distinguish between high efficiency phosphorylation of a PNA-peptide conjugate that hybridises poorly and low efficiency protein kinase modification with a higher concentration. Comparing the Cy3/fluorescein ratio of every feature allows relative quantification of the fluorescence, independent of any other factors, allowing Abl to be accurately profiled, with the highest ratio of Cy3/fluorescein corresponding to the peptide best recognised by Abl (Scheme 2.14).

It should be noted that experiments were carried out twice to avoid any false positives. To minimise cost issues, it was also possible to strip the probes and re-hybridise the 22,575 custom DNA microarrays up to three times, using a stripping buffer (2.5 mM Na_2HPO_4, 0.1% (v/v) SDS, 95 °C, 10 min).[161]



Scheme 2.14: Protein tyrosine kinase profiling using a PNA-encoded library and a secondary fluorescently labelled antibody

2.6 Data analysis

2.6.1 General concept

The dual colour system represented a two-channel microarray experiment, with the presence of two different fluorescent dyes, fluorescein and Cy3. Since fluorescein was used as a label for each library member, analysis of the array using the fluorescein channel confirmed that all library members had been synthesised and hybridised onto the DNA microarray. The second channel Cy3 was used to quantify the level of phosphorylation. After scanning the chip at two different wavelengths,

two images were obtained for each dye, and the signal ratio of Cy3/fluorescein was calculated for each peptide at each spot location.

2.6.2 General interpretation of the results

For the three screened kinases, data were reprocessed and interpreted in the same manner. The following explanations can thus be read as a general guideline. An Excel macro (see supplementary data on the attached DVD-ROM) allowed every DNA sequence to be correlated to its PNA tag and hence its peptide, which in combination with BlueFuse (BlueGnome, Cambridge, UK) allowed extensive mining and manipulation. BlueFuse works with a Bayesian framework using statistical models to produce a single confidence estimate for each ratio, where every duplicate is compared in terms of spot quality, signal intensity and regularity between the Cy3 and fluorescein channels. Values vary from 0 to 1, with the highest value corresponding to a high degree of confidence in the calculated ratio of Cy3/fluorescein. Although similarities with gene expression profiling exist, the analysis used here differed in many respects due to the fact that the majority of the features in the Cy3 channel had little if any signal (most peptides were non-phosphorylated). In order to gain valuable data, only features (specific peptides) with a Cy3 intensity twice that of the background were considered. Thus, a number of single initial signals (duplicates of each DNA oligomer existed scattered across the array) were obtained. This data was then verified by identification of all duplicates from within these single signals removing those with a standard deviation (StdDev) higher than 0.25 (comparing normalised ratios of Cy3/fluorescein intensities for each feature with values ranging from 0 to 1). Finally, the meaningful number of peptides phosphorylated by a specific tyrosine kinase ("hits") was given with very high confidence. Raw and reprocessed data for the three screened kinases are available as excel files on the attached DVD-ROM. After being reprocessed using BlueFuse, data were visualised as a 2D scatter plot using Spotfire (Cambridge, USA), with the y axis corresponding to the ratio of Cy3/fluorescein and the x axis to the amino acids present at one of the randomised positions (AA^1, AA^2, AA^3 and AA^4) of the library. Thus, profiles (percentages) of the amino acids that were the most accepted for each position were easily and quickly obtained, before being compared with sequences of

known kinase substrates. Moreover, since the highest ratios of Cy3/fluorescein correspond to the best peptides recognised by the tyrosine kinase, the top peptide sequences were generated. Finally, a protein basic local alignment search tool (BLAST) search using the Swiss-Prot database was performed with the top 25 peptide sequences.[162, 163] Analysis was run against the non-redundant protein sequence database. To have a biological meaning, only the highest scores of similarities as well as the lowest expectation values were considered. In fact, a low expectation value is generally associated with a significant match among the database sequences.[164] Thus known and putative interacting partners were identified for each profiled protein tyrosine kinase.

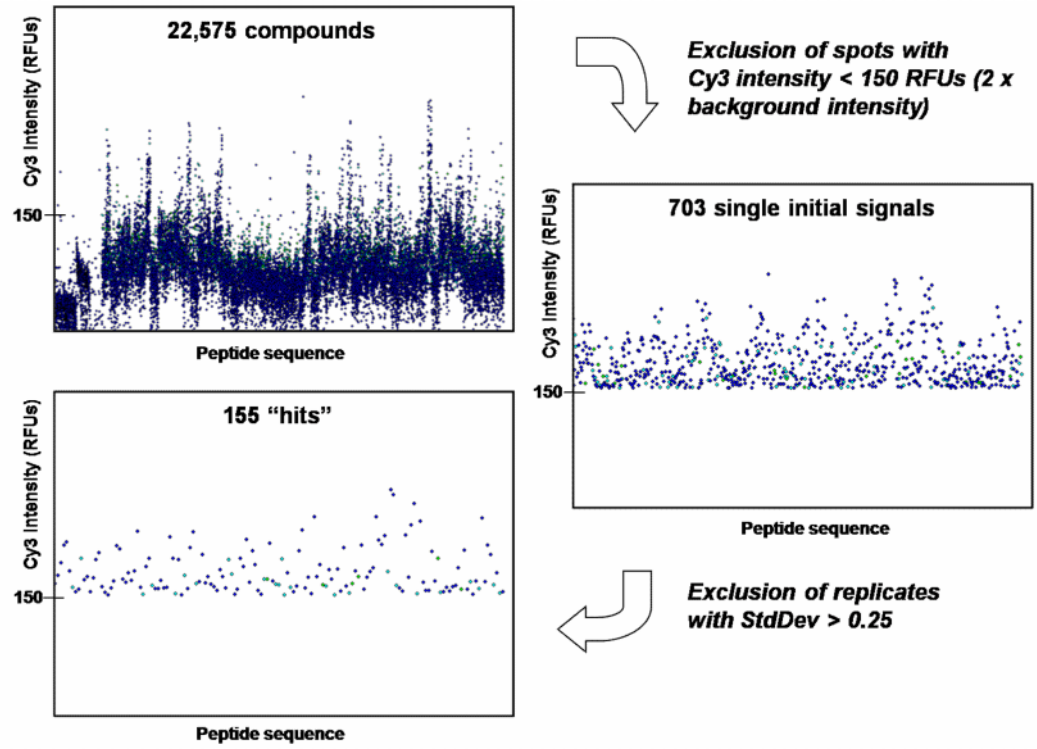
2.7 Enzymatic assays with Abelson tyrosine kinase (Abl)

2.7.1 Introduction

The non-receptor tyrosine kinase Abl is involved in the regulation of cell proliferation, transcription, apoptosis, cell adhesion and stress responses.[160, 165] A fusion protein called the breakpoint cluster region (Bcr)-Abl has been associated with human chronic myelogenous leukaemia (CML). The Bcr-Abl gene, which is formed by a translocation between chromosomes 9 (Abl gene) and 22 (Bcr gene), is overexpressed in cells compared to the normal gene that encodes the Abl kinase and is responsible for progression of the disease.[166] Patients suffering from CML are treated with the well-known Bcr-Abl kinase inhibitor imatinib **2.3**.[118] In the case of poor efficiency of imatinib **2.3**, a new class of inhibitors such as dasatinib and nilotinib that have recently emerged on the market are administrated.[167]

2.7.2 Specificity of Abl

Kinase assays were carried out using Abl kinase (mouse recombinant protein) and data were reprocessed as explained in section 2.6. From the 22,575 possible substrates, 703 single initial signals were firstly identified, resulting in 155 final peptides (“310 duplicative hits”) that were phosphorylated by Abl (Scheme 2.15). Sequences of the 155 “hits” recognised by Abl are listed in Appendix I.



Scheme 2.15: Scatter plots for Abl (reprocessed using BlueFuse); Cy3 intensity was expressed in relative fluorescence units (RFUs)

Broad specificity for the amino acids in positions AA^4 and AA^1 was found whereas high selectivity was clearly observed for the amino acids around the phosphorylation site (Tyr) in the AA^3 and AA^2 positions. Glu seems to play a crucial role in Abl phosphorylation since it was the most frequent amino acid found in each randomised position, confirming that this protein tyrosine kinase accepts acidic residues around the phosphorylation site (Asp was not chosen as a building block in the library synthesis due to its similarity to Glu).[132, 137] Glu (85 "hits" out of 155, 55%) was clearly the most accepted amino acid in the AA^3 position. Thus, an acidic, negatively charged amino acid on the amino side of tyrosine seems to be important for efficient phosphorylation by Abl. A smaller proportion of Val (18%) and Ser (11%) were also observed. In the AA^2 position, Abl displayed a preference for polar amino acids such as Ser (35%) and acidic residues such as Glu (32%). In the AA^4 and AA^1 positions, Abl was observed to be less specific and accepted a broad range of amino acids such as Ser (24%), Glu (19%), D-Val (17%) in the AA^4 position and Ser (22%), Val (18%), Glu (13%) in the AA^1 position (Figure 2.13).

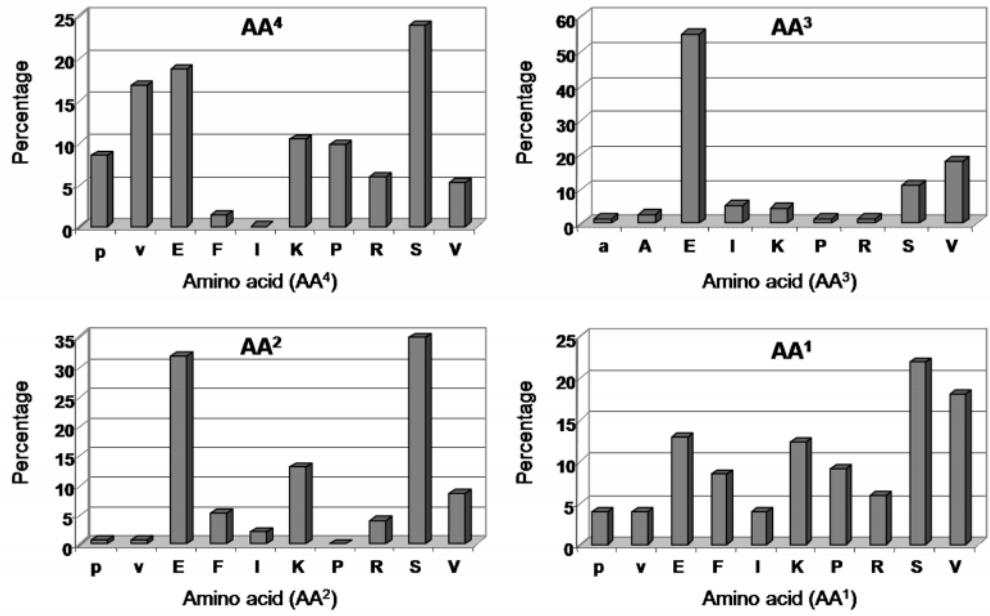


Figure 2.13: Bar graphs showing the proportion of amino acids at the four randomised positions in the 155 "hits"

When these results were compared with the sequences of known Abl kinase substrates such as tyrosine-protein phosphatase non-receptor type 6 (PTN6), mouse double minute 2 (Mdm2), docking protein 1 (dok1) and myoblast determination protein 1 (MYOD1), marked similarities were displayed. The preferred amino acids Glu (Asp), Ser and Val identified from the library screening were found around the phosphorylation site tyrosine, confirming the accuracy of the results (Table 2.2).

Table 2.2: Pentapeptide sequences of known Abl substrates (the known Abl substrate phosphorylation sites are indicated and amino acids in bold letters were identical to the amino acids identified at the AA^1, AA^2, AA^3 and AA^4 positions of the library)

Substrate	Swiss-Prot	Position	Sequence
PTN6	PTN6_HUMAN (MOUSE)	Y536	**SEY**GN
PTN6	PTN6_HUMAN (MOUSE)	Y564	**DVYE**N (**EVYE**N)
Mdm2	MDM2_HUMAN (MOUSE)	Y394 (Y393)	**EDYS**Q **(DDYS**Q)
Dok1	DOK1_MOUSE	Y314	**SVYSD**
Dok1	DOK1_MOUSE	Y397	**E**G**YE**L
MYOD1	MYOD1_MOUSE	Y30	**D**F**YDD**
MYOD1	MYOD1_MOUSE	Y212	M**DYS**G

2.7.3 Specificity of the top peptide sequences for Abl

The data from the 155 "hits" could also be analysed to generate the sequences of the best 10 substrates accepted by Abl (Table 2.3) using the ratios of Cy3/fluorescein calculated for each peptide (Figure 2.14). In this study, the optimal peptide (with the highest ratio of Cy3/fluorescein) recognised by Abl was FQSEYEVIK. This peptide contained the SEY motif that appeared 5 times in the top 10 peptides, revealing this to be an important factor in Abl specificity. Moreover, among the 155 "hits", 21 peptides contained the SEY motif and 14 were found in the top 40 peptide sequences (see Appendix I). By generating the top peptide sequences, it was also possible to identify amino acids that were not necessary present in a large number at AA^1, AA^2, AA^3 and AA^4 positions but had a high ratio of Cy3/fluorescein and consequently an important biological meaning. For example, 8 peptides contained Ile (5%, Figure 2.13) in the AA^3 position and these peptides were among those giving the highest ratios of Cy3/fluorescein (Figure 2.14). In the top 10 peptide sequences, Ile appeared 3 times in the AA^3 position (Table 2.3). The Abl optimal substrate predicted by Songyang[132] contains the motif I/VYXXP, where X represents any amino acid. Although Pro was replaced by Ile in the library (Figure 2.10), Ile was successfully identified as an important amino acid at the AA^3 position. It is known that Abl accepts a wide range of substrates[168] and one of the advantages of the method was that optimal peptides for Abl, rather than consensus sequences, were identified. Moreover, unlike other techniques[137] that gave large number of "hits" without relative quantification, the approach described here offered the advantage of not only detecting phosphorylated peptides but also quantifying the extent of phosphorylation (the higher the ratio of Cy3/fluorescein, the greater the extent of phosphorylation).

Table 2.3: Sequences of the top 10 peptides

Peptide sequence	Ratio Cy3/fluorescein	Peptide sequence	Ratio Cy3/fluorescein
1) FQSEYEVIK	4.50	6) FQSIYEPIK	3.44
2) FQSEYESIK	4.45	7) FQSIYSPIK	3.32
3) FQSEYSFIK	4.16	8) FQPEYSEIK	3.29
4) FQSEYVFIK	4.03	9) FQSEYEEIK	3.27
5) FQVIYESIK	3.62	10) FQVEYESIK	3.26

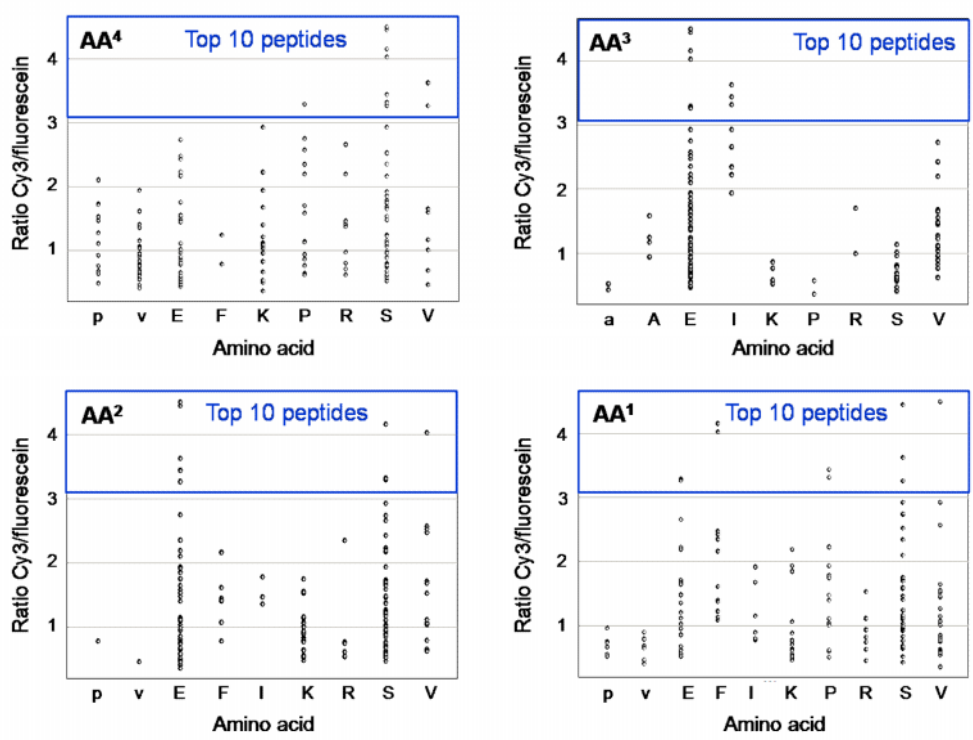


Figure 2.14: Representation of the 155 "hits" recognised by Abl

2.7.4 Interacting partners for Abl

Only the most important amino acids around the phosphorylation site (pentapeptide) were taken into account in the BLAST search,[162] and analysis was run against the mouse organism. Among the top 25 peptides (see Appendix I), proteins that are involved in known Abl signalling pathways[169, 170] and implicated in leukaemia[171] were readily identified (Table 2.4).[160] In particular, the second most phosphorylated peptide (SEYES) was the well known Abl substrate, c-Jun amino-terminal kinase (JNK).[169] In response to DNA damage, JNK phosphorylation of 14-3-3 proteins induces dissociation of the complex formed between the cytoplasmic Abl and the 14-3-3 proteins, resulting in translocation of Abl to the nucleus and DNA-damage-induced apoptosis.[169] Chemokines and their receptors control leukocyte migration and maturation, lymphocyte traffic, and C-C chemokine receptors have been found to be expressed in B cells of different types of leukaemia.[171] It has recently been reported that Abl phosphorylated the epidermal growth factor receptor (EGFR) on a specific tyrosine phosphorylation site and inhibited endocytosis of the EGFR. Abl also increased cell-surface expression of EGFR and Abl/EGFR signalling might be

involved in human tumours.[170] Although the role of Abl with G protein coupled receptor (GPCR) kinases and diacylglycerol kinase gamma is not fully understood, our data revealed that Abl might interact with these proteins.[172, 173] GPCRs are the largest family of cell-surface receptors involved in a diverse range of signalling pathways, implicated in numerous disease states (*e.g.* cancer, cardiovascular disease, neurodegenerative disorders),[174] and Abl may play an important role in GPCR signalling.[172] So far it has been shown that overexpression of diacylglycerol kinase gamma inhibited macrophage differentiation of human leukaemia cells.[173]

Table 2.4: Predicted proteins

Protein name	Swiss-Prot	Identified "hit"
G-protein coupled receptor kinase	GRK6_MOUSE	1) SEYEV
c-Jun amino-terminal kinase	JIP2_MOUSE	2) SEYES
Diacylglycerol kinase gamma	DGKG_MOUSE	11) SEYSS
G-protein coupled receptor kinase	GRK5_MOUSE	14) EVYSS
C-C chemokine receptor	CCR6_MOUSE	24) RVYSE
Epidermal growth factor receptor substrate	EP15_MOUSE	25) PVYEK

To illustrate the generalisation of this approach, two other tyrosine kinases, human epidermal growth factor 2 (Her2) and vascular endothelial growth factor receptor 2 (VEGFR2) were profiled.

2.8 Enzymatic assays with human epidermal growth factor 2 (Her2)

2.8.1 Introduction

Her2 (also known as Her2/neu, erbB2 and EGFR2) belongs to the EGFR family of receptor tyrosine kinases involved in signal transduction pathways (cell growth and differentiation).[175] Overexpression (amplification) of Her2 has been found in breast, ovarian, gastric, lung and colon cancers,[112] with the Her2 gene amplified in more than 25% of human primary breast cancers.[176] The humanised monoclonal antibody trastuzumab can block proliferation of cells overexpressing Her2, increasing survival of patients suffering from breast cancer, when used in combination with chemotherapy.[177] A second-generation of inhibitors[120] such as lapatinib that is able to

block activity of both EGFR and Her2 have been used to treat advanced breast cancer.[178]

2.8.2 Specificity of Her2

Kinase assays were carried out using Her2 kinase (human, recombinant, *N*-terminal GST tag) and data were reprocessed as explained in section 2.6. Some 2879 single initial signals were found, giving 299 final peptides (598 duplicates) that were selected as "hits" for Her2. Sequences of the 299 "hits" recognised by Her2 are listed in Appendix II. Profiles of the amino acids that were the most accepted for each randomised position are shown in Figure 2.15.

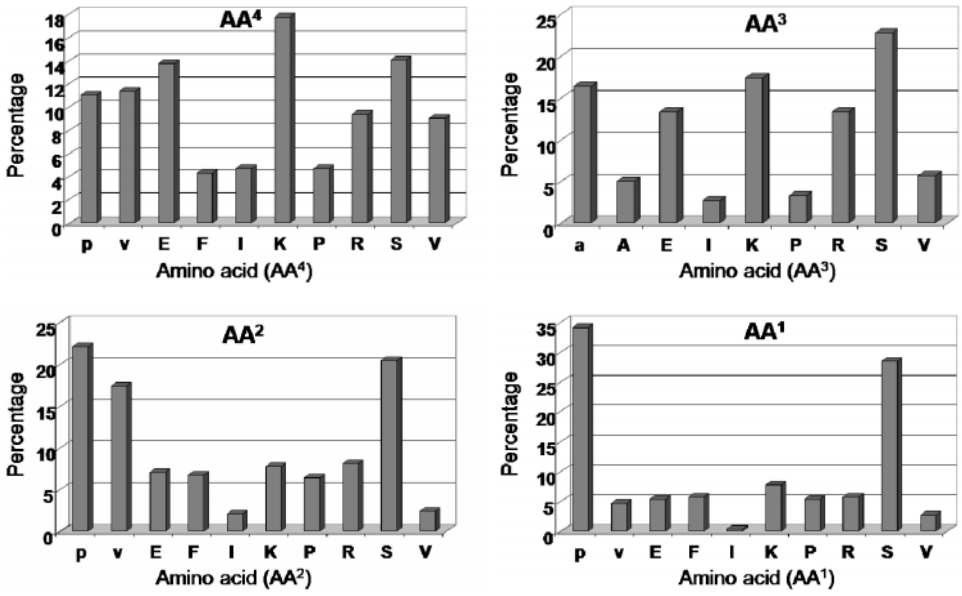


Figure 2.15: Bar graphs showing the proportion of amino acids at the four randomised positions in the 299 "hits"

Surprisingly, Her2 showed a remarkable preference for the unnatural amino acid D-Pro at the *C*-terminus site of the phosphorylation site. Thus, looking at the 299 "hits", in the AA^1 position, D-Pro was the most accepted amino acid (34%) along with Ser (28%). This was also observed in the AA^2 position, where D-Pro (22%), Ser (20%) and D-Val (17%) were the most common amino acids. Moreover, in the top 15 peptides (appendix II), 10 peptides in the AA^1 position and 7 in the AA^2 position contained D-Pro, with the sequence D-Pro/D-Pro (AA^2/AA^1) represented five times.

This preference was also confirmed in Figure 2.16, where the highest ratios of Cy3/fluorescein in positions AA^2 and AA^1 were peptides containing D-Pro. These results were particularly interesting since even though L-Pro was present in positions AA^1 and AA^2, Her2 had a clear preference for D-Pro. The marked preference for this unnatural amino acid was unexpected and was not observed in other library screens. However, it is already known that protein kinase C can phosphorylate both D and L-amino acids.[179] In this study, several nonapeptides (Lys-Arg-Pro-Ser-Gln-Arg-Ala-Lys-Tyr) containing one, two or nine D-amino acids at different positions were tested. It was found that protein kinase C phosphorylated all the peptides, even peptides containing only D-amino acids or a D-serine, at different kinetic rates. The Michaelis constant K_m values for the peptides containing one D-amino acid were similar to the one obtained with L-amino acids, but the presence of a higher number of D-amino acids or the replacement of serine by threonine resulted in a lower phosphorylation rate and an increased K_m value.

Her2 was less specific for the amino acids in positions AA^4 and AA^3 but a small preference for Lys (18%), Glu (14%) and Ser (14%) in the AA^4 position and Ser (23%), Lys (17%) and D-Ala (16%) in the AA^3 position could be noticed (Figure 2.15). Since the most preferred amino acids at the randomised positions were unnatural amino acids, it was not possible to compare these results with known Her2 substrates.[180] It should be noted that low ratios of Cy3/fluorescein were obtained due to very bright fluorescent spots observed when the chip was scanned using a fluorescein filter (Table 2.5) but the experiments were run twice, giving similar results regarding Her2 specificity (see section 2.8.3).

Table 2.5: Sequences of the top 15 peptides

Peptide sequence	Ratio Cy3/fluorescein	Peptide sequence	Ratio Cy3/fluorescein
1) FQpRYPpIK	1.40	9) FQEAYRSIK	1.09
2) FQVRYFpIK	1.34	10) FQIaYppIK	1.08
3) FQFaYppIK	1.32	11) FQPRYSpIK	1.08
4) FQpIYppIK	1.20	12) FQISYppIK	1.07
5) FQpAYpRIK	1.16	13) FQRSYFpIK	1.06
6) FQKIYppIK	1.14	14) FQVRYRSIK	1.06
7) FQFAYpSIK	1.11	15) FQSIYRSIK	1.03
8) FQPAYFpIK	1.09		

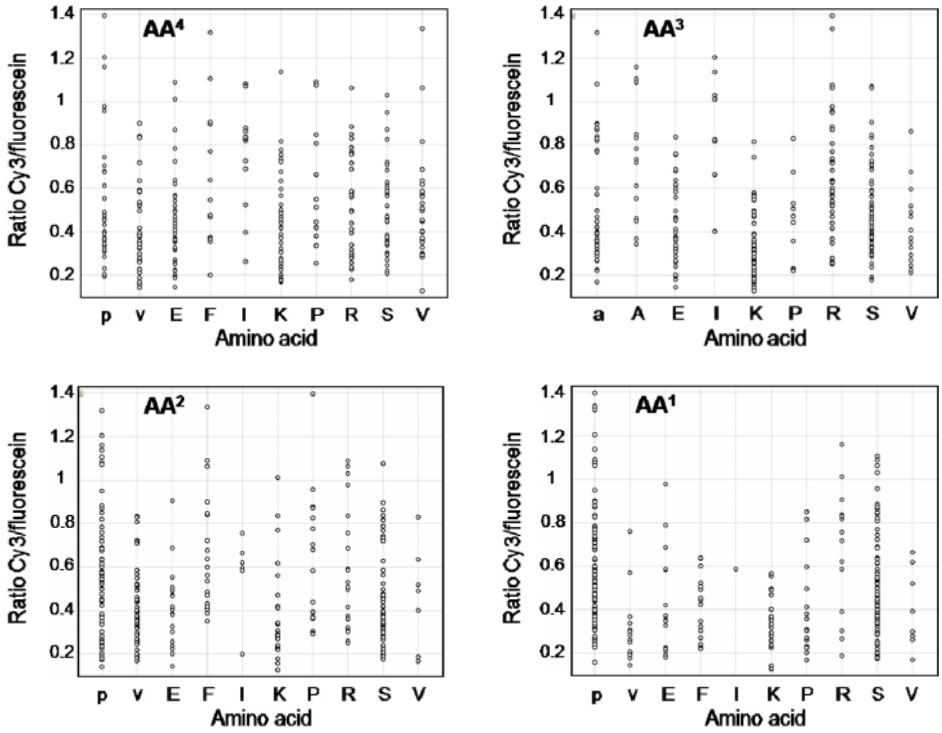


Figure 2.16: Representation of the 299 "hits" recognised by Her2

2.8.3 Interacting partners for Her2

Among the top 25 peptides (see Appendix II), six peptide sequences contained natural amino acids so that the BLAST search was run against the human organism using these six peptides. To have robust data, it was necessary to enter the whole peptide sequence -Phe-Gln-AA^4-AA^3-Tyr-AA^2-AA^1-Ile-Lys- instead of considering only the 4 amino acids present around the phosphorylation site. By using the pentapeptide sequences, low similarity scores and high expectation values were obtained, and consequently the identified proteins did not have a significant biological meaning. Although only six peptide sequences were used for the BLAST search, proteins that are present in breast cancer[181, 182] or known to interact with Her2[183] were identified (Table 2.6). For instance, the breast tumour-amplified kinase gene that encodes a putative serine/threonine kinase has been found to be amplified and overexpressed in human breast cancer cell lines.[181] The role of Her2 in mitogen-activated protein kinase activation is vital and Her2 is used as a coreceptor for GPCR signalling in the heart.[183] The ephrin receptors are the largest group of human receptor tyrosine kinases and play an important role as regulators in the development

and the progression of cancer. It has been demonstrated that upregulation of the ephrin receptor EPHA7 induces tumorigenesis and invasiveness of different types of breast cancer cells.[182]

Table 2.6: Predicted proteins (in the identified "hits", amino acids in bold letters were identical to the protein sequence amino acids)

Protein name (sequence)	Swiss-Prot	Identified "hit"
Breast tumour-amplified kinase (YQETYKRI)	STK6_HUMAN	16) F**QE**I**YKRI**K
G protein coupled receptor 158 (EIYKRKK)	GP158_ HUMAN	16) FQ**EIYKRI**K
Ephrin receptor EPHA7 (FQTRYPS)	EPHA7_ HUMAN	25) **FQ**S**RYPS**IK

2.9 Enzymatic assays with vascular endothelial growth factor receptor 2 (VEGFR2)

2.9.1 Introduction

The receptor tyrosine kinase VEGFR2 (also known as the kinase insert domain-containing receptor (KDR) and Flk1) is a member of the vascular endothelial growth factor (VEGF) family that plays crucial roles in angiogenesis and vasculogenesis.[184] Among the VEGFR inhibitors that have been developed, one of the most efficient is bevacizumab, a humanised monoclonal antibody that binds VEGF before its attachment to its receptors. It improves survival of patients with non-small cell lung, colon, renal and breast cancers but has to be used in combination with chemotherapy.[185] Alternatively, small-molecule inhibitors that target several receptor tyrosine kinases including VEGFR2 have appeared. Thus, sunitinib maleate[186] has been approved for treatment of gastrointestinal stromal tumours and advanced renal-cell carcinoma.[187] Another example is sorafenib that is also used for treatment of kidney cancer as well as advanced primary liver cancer.[188]

2.9.2 Specificity of VEGFR2

Kinase assays were carried out using VEGFR2 kinase (human, recombinant, *N*-terminal GST tag) and data were reprocessed as explained in section 2.6. Some 8156 single initial signals were found. Thus, unlike the two other kinases, VEGFR2 was much less specific and accepted a broad range of amino acids, giving a much greater

number of 829 "hits" (1658 duplicate peptides) (sequences of the 829 "hits" recognised by VEGFR2 are available as excel files on the attached DVD-ROM).

A slight preference for polar amino acids (both acidic and basic) immediately adjacent to the phosphorylation site (positions AA^3 and AA^2) could be observed: Glu (AA^3/AA^2:20%/16%), Lys (AA^3/AA^2:17%/16%) and Ser (AA^3/AA^2:17%/14%) were the most commonly observed amino acids. In the AA^4 and AA^1 positions, Lys (15%) and D-Pro (18%) were respectively the most common amino acids (Figure 2.17).

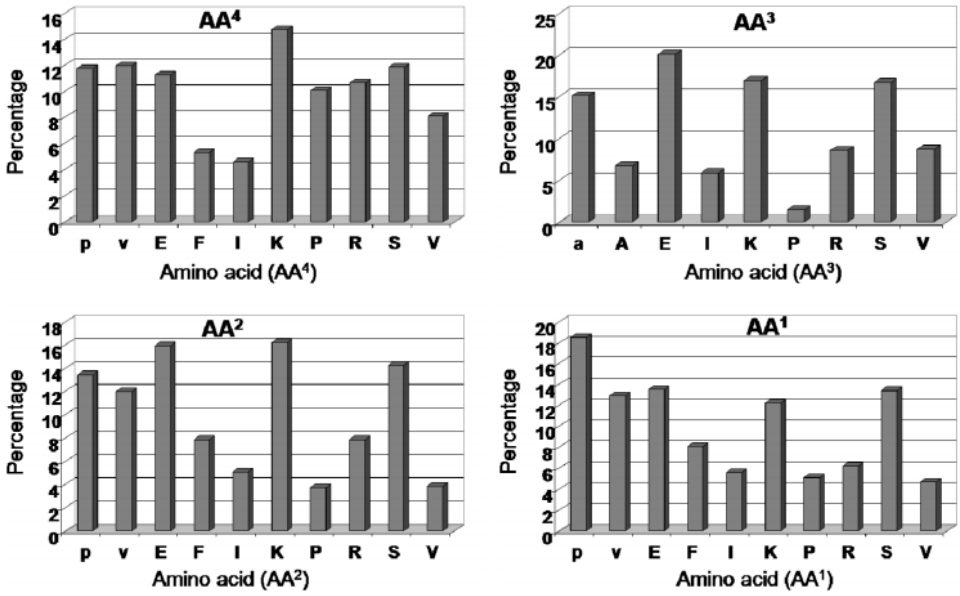


Figure 2.17: Bar graphs showing the proportion of amino acids at the four randomised positions in the 829 "hits"

2.9.3 Specificity of the top peptide sequences for VEGFR2

Although Ile (5%) was the least accepted amino acid in the AA^4 position (Figure 2.17), peptides containing Ile at this specific position were the ones with the highest ratios of Cy3/fluorescein (Figure 2.18). This was confirmed with the number of peptide sequences containing the hydrophobic amino acid Ile in the AA^4 position: 5 and 14 for the top 10 (Table 2.7) and 50 peptide sequences (see Appendix III) respectively. This observation has not been reported in the literature yet. Moreover, since the motifs IAY and IAYP appeared 3 times and twice in the top 10 and 3 "hits" respectively, it seems to be important in VEGFR2 specificity (Table 2.7).

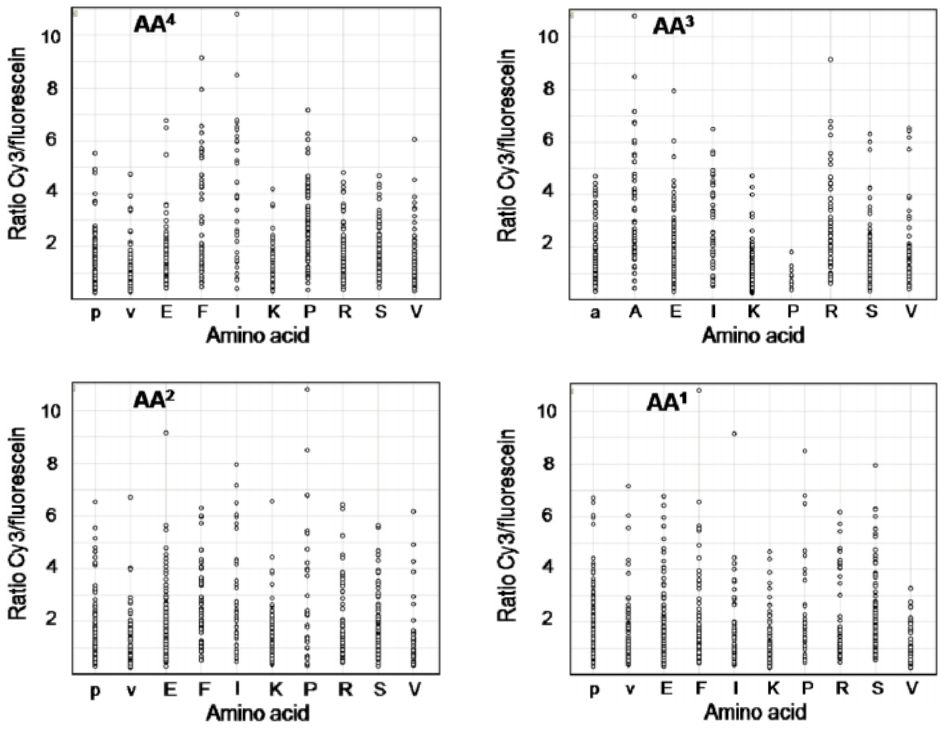


Figure 2.18: Representation of the 829 "hits" recognised by VEGFR2

Table 2.7: Sequences of the top 10 peptides

Peptide sequence	Ratio Cy3/fluorescein	Peptide sequence	Ratio Cy3/fluorescein
1) FQIAYPFIK	10.80	6) FQIRYPPIK	6.79
2) FQFRYEIIK	9.15	7) FQEAYPEIK	6.77
3) FQIAYPPIK	8.50	8) FQIAYvpIK	6.71
4) FQFEYISIK	7.95	9) FQFRYKFIK	6.56
5) FQPAYIvIK	7.17	10) FQIVYppIK	6.53

2.9.4 Interacting partners for VEGFR2

Similarly to the two other screened tyrosine kinases, a BLAST search was run against the human organism. To obtain high scores of similarities and low expectation values, both pentapeptide sequences and the whole peptide sequence -Phe-Gln-AA^4-AA^3-Tyr-AA^2-AA^1-Ile-Lys- were used for the BLAST search. Results were also in agreement with known VEGFR2 signalling proteins (Table 2.8).[189-191] For example, it has been reported that leptin and its receptors favour the development of breast cancers, increase the levels of VEGFR and the expression of VEGFR2 in tumour cells.[189] The protein kinase C controls the down-regulation of VEGFR2 *via*

potential phosphorylation of serine residues in the VEGFR2 carboxyl-terminal domain.[190] Sphingosine 1-phosphate induces endothelial cell proliferation and migration *via* endothelial differentiation GPCRs. A signalling pathway involving GPCRs, VEGFR2, the tyrosine kinase Src and the CrkII adaptor protein has been shown to improve the mobility of human umbilical vein endothelial cells.[191]

Table 2.8: Predicted proteins (in the identified "hits", amino acids in bold letters were identical to the protein sequence amino acids)

Protein name (sequence)	Swiss-Prot	Identified "hit"
Leptin receptor isoform 1 (FQIRY)	LEPR_HUMAN	6) **FQIRY**PPIK
Protein kinase C, theta (IVYRD)	KPCT_ HUMAN	12) **IVYRE**
Endothelial differentiation G protein coupled receptor 6 (IIYSF)	EDG6_ HUMAN	23) **IIYSF**

2.10 Conclusion

The high-throughput screening of 10,000 tyrosine kinase Abl substrates in a single experiment has been demonstrated *via* the detection of tyrosine-phosphorylated peptides ("hits") using a non radioactive, selective and sensitive method based on a primary anti-phosphotyrosine antibody and a fluorescently labelled secondary antibody. The use of PNA-encoded libraries simplified the deconvolution process and the microarray format allowed minimal amounts of kinase (60 pmol), library (5 nmol) and antibodies (0.4 μg per antibody) to be used. ALL 10,000 members of the library were analysed, allowing a unique insight into the substrate requirements of Abl; however it is a general method as there is no need to have any previous knowledge about an optimal kinase substrate. The dual colour approach (ratio of Cy3/fluorescein) was essential to allow accurate profiling, and gave a set of peptides that were phosphorylated in agreement with known Abl interacting partners, while new target proteins were quickly and simply identified by analysis of the top peptide sequences and BLAST searching. The approach was used to profile and identify specific proteins for two other tyrosine kinases, Her2 and VEGFR2, showing its generality, while adaptations[140] will allow serine and threonine kinases to be analysed allowing rapid and efficient substrate deorphaning.

Chapter 3: PNA-encoded positional scanning libraries to profile protease substrate specificity

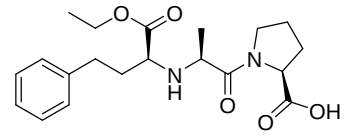
3.1 Proteases

Proteases, also known as proteinases, peptidases or proteolytic enzymes, specifically catalyse the hydrolysis of amide bonds in peptides and proteins. Proteases constitute one of the largest eukaryotic protein families, with more than 550 proteases identified in the human genome (around 2% of the gene content).[192, 193] Proteases are classified into six categories: serine (*e.g.* chymotrypsin, trypsin, subtilisin), cysteine (*e.g.* papain, caspase), threonine (*e.g.* proteasome), aspartic acid (*e.g.* renin, pepsin, human immunodeficiency virus (HIV)-1 protease), metallo- (*e.g.* thermolysin, carboxypeptidase A) and glutamic acid proteases. Proteases can cleave either at the *N*-terminus (so-called aminopeptidases) or *C*-terminus (so-called carboxypeptidases), or within the peptide/protein (so-called endopeptidases). For aspartic acid, metallo- and glutamic acid proteases, a molecule of water acts as nucleophile to hydrolyse peptide bonds, with the amide bond polarised by Lewis acid catalysis. In the case of serine, cysteine and threonine proteases, the mechanism usually involves a catalytic triad, where a histidine residue is used to activate a serine, cysteine or threonine side chain and place it in a position that favors deprotonation. This results in the formation of an “alkoxide” or “thiolate”, which are better nucleophiles than alcohols and thiols.[193]

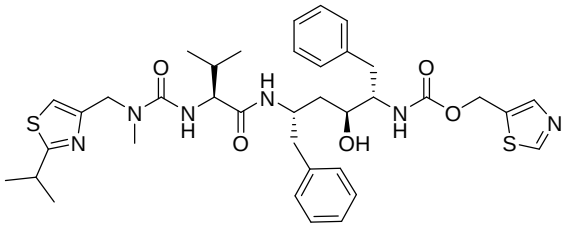
Proteases play a multitude of fundamental roles in many biological pathways such as cell or viral proliferation, differentiation, apoptosis, DNA replication and transcription, blood coagulation, tissue remodelling, wound repair, inflammation and angiogenesis.[193] Consequently, aberrant regulation of protease activity results in important diseases such as viral infections, cancer, Alzheimer’s, inflammatory, renal and cardiovascular diseases.[194] Proteases are thus significant drug targets in the pharmaceutical area, and many protease inhibitors have been commercialised.

3.2 Protease inhibitors

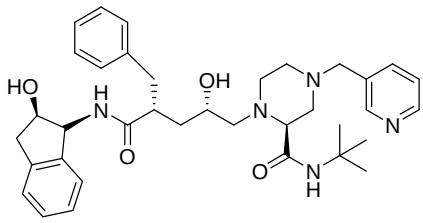
Several important drugs in the market or in clinical trials are protease inhibitors.[194, 195] One of the first major protease inhibitors are those of the metalloprotease angiotensin-converting enzyme that are used to regulate blood pressure and for the treatment of post myocardial infarction and congestive heart failure. An example is the pro-drug enalapril **3.1** (Figure 3.1), which is metabolised *in vivo* to the active form enalaprilat by hydrolysis of the ethyl ester following oral administration. Several antiretroviral drugs have also been developed, for instance ritonavir **3.2** and indinavir **3.3** are inhibitors of the HIV protease and are used in the treatment of acquired immunodeficiency syndrome (Figure 3.1). Proteases and their inhibitors can also be used as diagnostic and prognostic biomarkers. The serine protease plasminogen activator and the plasminogen activator inhibitor-1 as well as the serine protease kallikrein 3 have been successfully used as diagnostic biomarkers in breast and prostate cancers respectively.[195] So far effective protease inhibitors have been mainly applied to treat cardiovascular, inflammatory, kidney and infectious diseases but many other therapeutic areas still need to be explored. To design efficient drugs against proteases and identify new substrates, it is necessary to understand the biological functions and the substrate specificity of proteases.



Enalapril 3.1



Ritonavir 3.2



Indinavir 3.3

Figure 3.1: Chemical structures of representative protease inhibitors

3.3 Protease substrate specificity

3.3.1 Introduction

Proteases are able to selectively recognise potential targets among a large number of substrates. This key feature called substrate specificity allows for maintenance of the correct functioning of most biological pathways. The knowledge of the substrate specificity of a protease is vital to understand its function, and to identify its physiological substrates and elaborate selective inhibitors. The standard nomenclature for protease substrate cleavage is shown in Figure 3.2. As a general rule, amino acids on the *N*-terminal side of the scissile bond (the cleavable bond) are labelled P_1, P_2, P_3, etc. ("non-prime side"), while amino acids on the *C*-terminal side of the scissile bond are labelled P_1', P_2', P_3', etc. ("prime side"). The corresponding binding sites are referred to as S_1, S_2, S_3, S_1', S_2', S_3', etc. The amino acids in the peptide immediately adjacent to the cleavage site (P_1, P_1') but also the ones in the other positions (P_2, P_2', P_3, P_3', etc.) play a significant role in protease specificity. A number of methods to profile protease substrate specificity have been developed by identifying the preferred "non-prime" and "prime" amino acids. The most common strategies that are based on the use of irreversible inhibitors, Edman degradation of peptide mixtures, positional scanning libraries of fluorogenic substrates, microarrays and FRET-based peptide libraries are precisely explained below.

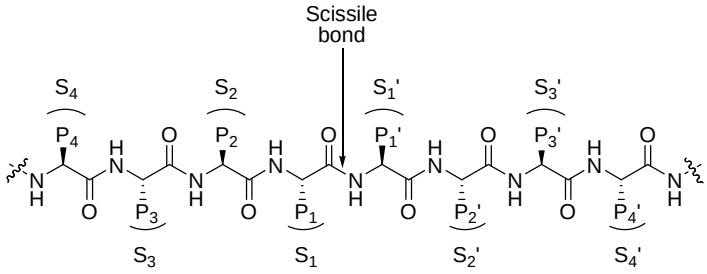
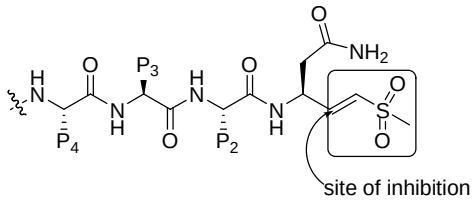


Figure 3.2: Standard nomenclature for protease substrate cleavage

3.3.2 Use of irreversible inhibitors

The "non-prime side" substrate specificity of the proteasome has been studied using positional scanning libraries of peptide vinyl sulfones (Scheme 3.1).[196] Peptide vinyl sulfones acted both as inhibitors of the proteasome and substrate mimics that were

covalently bound to the protease active site *via* Michael addition of the active site nucleophile onto the vinyl sulfone. To determine inhibitor binding, positional scanning libraries or individual compounds were incubated with the protease in a competition assay, followed by subsequent addition of a general radiolabelled inhibitor to label any remaining active site nucleophiles of the protease. Following separation of the protease with sodium dodecylsulfate polyacrylamide gel electrophoresis, the efficiency of the peptide vinyl sulfone as an inhibitor was assessed by the decrease in radioactivity in comparison with a 100% radiolabelled inhibitor used as a standard.



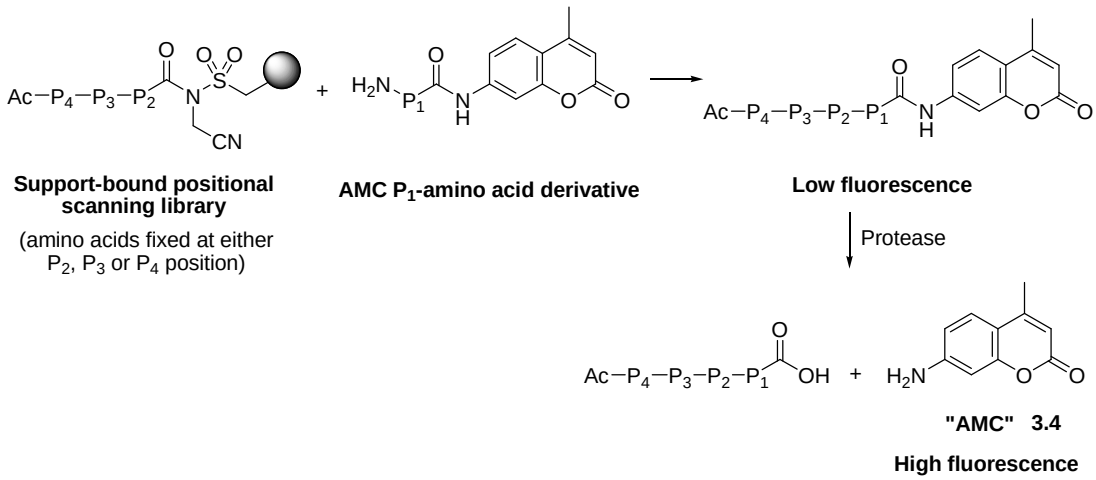
Scheme 3.1: General structure of a positional scanning library of P_1-asparagine peptide vinyl sulfones

3.3.3 Edman degradation of peptide mixtures

Another strategy, using mixtures of acetyl-capped peptides, has been used to profile the "prime side" substrate specificity of a cysteine protease.[197] The peptides were incubated with the protease, resulting in the formation of cleaved peptides containing a free amino group, whose sequences were determined by Edman degradation,[133] with the first round of sequencing giving information about the P_1' position, the second round on the P_2' position, etc. This approach was successfully applied to determine the substrate specificity of six enzymes belonging to the matrix metalloprotease family.[198] The "prime side" substrate specificity was obtained using a combinatorial peptide library that was subjected to partial proteolytic cleavage (up to 10% cleavage) and subsequent Edman degradation. The results obtained from the first round of sequencing were then used to design a second library, where preferred amino acids were fixed on the *C*-terminal side, providing information about the "non-prime side" specificity of the six proteases.

3.3.4 Positional scanning libraries of fluorogenic substrates

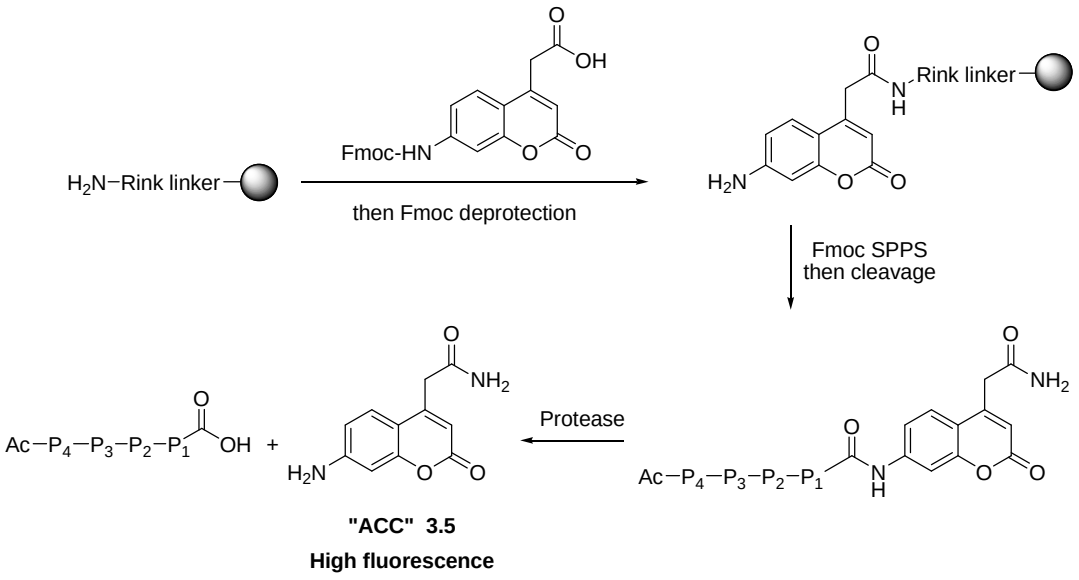
Positional scanning libraries of fluorogenic substrates have been used to rapidly determine the *N*-terminal ("non-prime side") substrate specificity of proteases. For example, Ellman profiled two serine proteases plasmin and thrombin using a tetrapeptide positional scanning library of fluorogenic 7-amino-4-methylcoumarin (AMC) peptide substrates.[199] In this strategy, a specific amino acid known to have affinities with the target protease had to be held constant at the P_1 position in order to direct the cleavage site. Since serine proteases require basic amino acids at the P_1 position, lysine was chosen as the fixed P_1-amino acid. The 6,859-member library was composed of three sub-libraries, in which one position (P_2, P_3 or P_4) was respectively fixed, while the other two contained an equimolar mixture of 19 amino acids. The three sub-libraries were synthesised by solid-phase synthesis using the *N*,*N*-cyanomethylacylsulfonamide "safety-catch" linker. Nucleophilic attack by the fluorogenic AMC lysine (P_1)-amino acid to the support-bound positional scanning library released compounds in solution. The products were added to 57 individual microwells, each one corresponding to an addressed position (one of the 19 amino acids fixed at either P_2, P_3 or P_4 positions) and containing 361 (19^2) AMC-derivatised tetrapeptides. Upon proteolytic cleavage at the P_1 position, the fluorophore AMC **3.4** was liberated in solution and its fluorescence quantified, allowing the rapid measurement of cleavage rates for a library of compounds (Scheme 3.2). This approach could be generalised to other classes of proteases since any amino acids could be fixed at the P_1 position. It represented an advantage over previous methods, where P_1-amino acid AMC derivatives had to be linked to the solid support through the side chain functional group, thus limiting the choice of amino acids to specific difunctional residues (*e.g.* aspartic acid).[200]



Scheme 3.2: General concept for the synthesis of a fluorogenic AMC positional scanning library

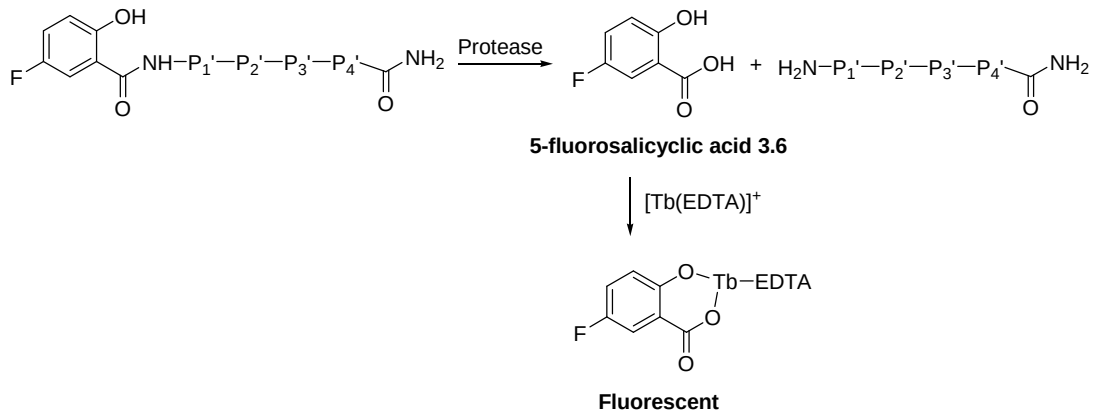
However knowledge about the protease substrate specificity at the P_1 position was required to synthesise the above P_1-fixed positional scanning library. To overcome this limitation, Ellman developed the bifunctional fluorophore 7-amino-4-carbamoylmethylcoumarin (ACC) **3.5** that allowed for the direct coupling of the fluorophore to a solid support and Fmoc-based solid phase peptide synthesis (SPPS) (Scheme 3.3).[201] A tetrapeptide positional scanning library of fluorogenic ACC peptide substrates composed of 137,180 members was used to study the P_1 position and profile the substrate specificity of several serine and cysteine proteases using a microtitre plate format. In that case, the P_1 position was diversified using 20 amino acids and the P_2, P_3 and P_4 positions contained an equimolar mixture of 19 amino acids. This P_1-positional scanning library was made up of 20 wells (one for every constant amino acid at the P_1 position), each containing 6,859 (19^3) peptide substrates. Similar kinetic profiles were obtained using a fluorogenic ACC or AMC P_1-fixed (lysine) positional scanning library. Moreover, the ACC fluorescence was found to be three-fold superior to AMC, thus reducing necessary enzyme and substrate concentrations.[201] Data obtained from the screening of a tetrapeptide positional scanning library of fluorogenic ACC peptides and a 4,000-member PNA-encoded small molecule library[202] were used to profile proteolytic activities in dust mite extracts and determine the cysteine protease activity of the house dust mite *Dermatophagoides pteronyssinus* allergen Der p 1.[203] Recently, a 160,000-member tetrapeptide positional scanning library of ACC peptide substrates with completely

randomised P_1, P_2, P_3 and P_4 positions was used to determine the *N*-terminal ("non-prime side") substrate specificity of papain-like cysteine and parasite proteases.[204] However it should be noted that this method will not detect cleavage between amino acids not attached to the fluorophore. In addition it can only look at half of the cleavage site.



Scheme 3.3: General concept for the synthesis of a fluorogenic ACC positional scanning library

A positional scanning library of fluorogenic peptides capped with 5-fluorosalicyclic acid **3.6** at the *N*-terminus was synthesised to study the "prime side" substrate specificity of bovine α-chymotrypsin.[205] Upon proteolytic cleavage, 5-fluorosalicyclic acid **3.6** was released in solution as a free carboxylic acid and coordinated to a terbium ion to generate a fluorescent signal (Scheme 3.4).

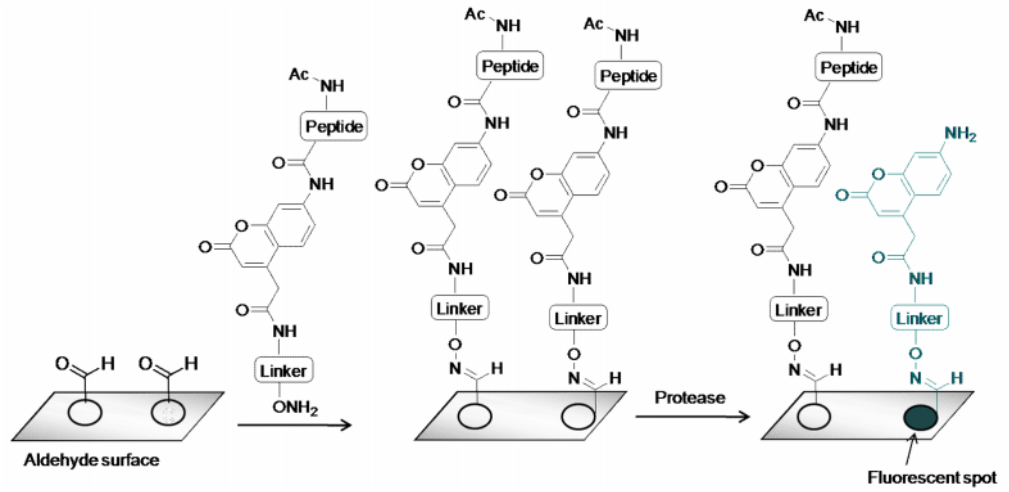


Scheme 3.4: Use of a fluorogenic positional scanning library to study the "prime side" protease substrate specificity

3.3.5 Microarray-based techniques

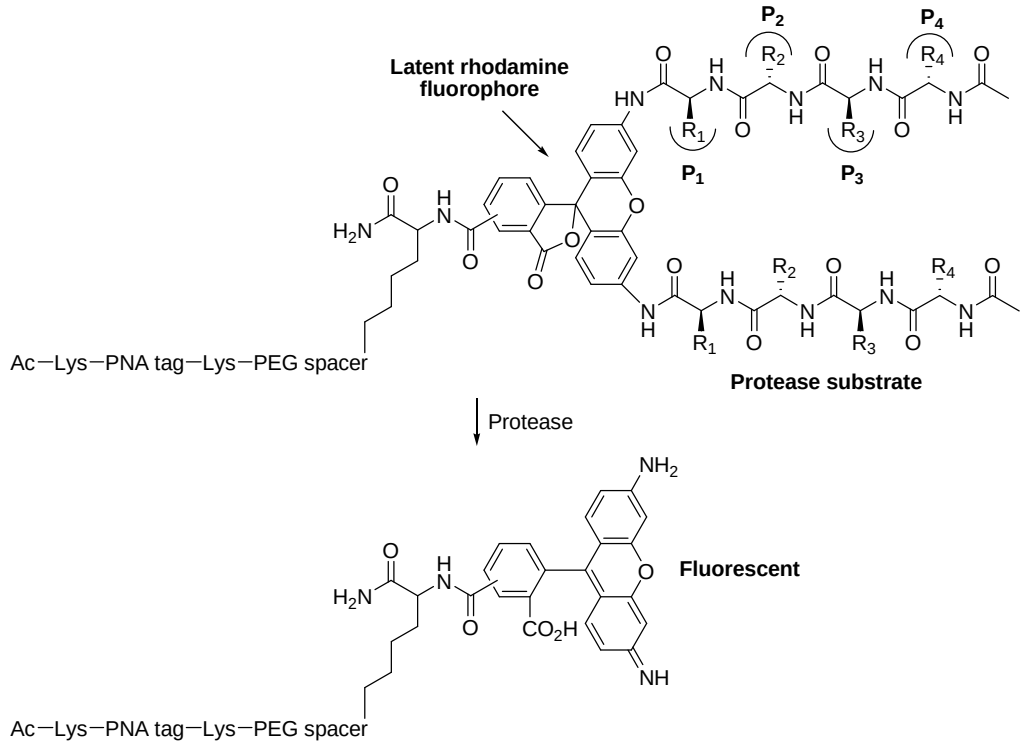
To increase the throughput of positional scanning libraries of coumarin-based substrates, identify subsite interactions between amino acids, and reduce the needed amounts of enzyme and library, Salisbury reported the use of alkoxylamine-derivatised fluorogenic ACC peptide microarrays to rapidly determine protease substrate specificity (Scheme 3.5).[105] Peptides with the general structure Ac-peptide-ACC-linker-ONH_2 were covalently linked to an aldehyde-derivatised microarray surface *via* oxime formation. The serine protease thrombin was profiled using a 361-member peptide positional scanning library (Ac-Ala-P_3-P_2-Lys-ACC-linker) and to ensure a good access of the immobilised peptides to the protease, the aldehyde slides were derivatised with BSA. The enzymatic assays were carried out by adding a protease solution directly on the whole microarray surface, with proteolytic cleavage identified *via* the generation of fluorescent spots. The results were in accordance with previous solution-phase assay data using positional scanning libraries of fluorogenic substrates.[199] Later on, this approach was adapted and nanodroplet microarrays were developed to profile 4 human blood serine proteases,[206] 11 papain-like cysteine and 13 serine proteases[207] using 722-member positional scanning libraries of ACC peptides. The substrates were printed as glycerol nanodroplets on a microarray, followed by aerosol addition of a protease solution to simultaneously initiate all enzymatic reactions. The advantages of the method were that there was no need for peptide and slide derivatisation and the substrates were kept in a low-volatile

solution-phase environment (glycerol), thus reducing non-specific binding, protease denaturation and accessibility issues observed when the substrates are covalently attached to the microarray. In a similar approach, Angenendt performed solution-based inhibition assays of Cathepsin D by printing the fluorogenic substrate, the enzyme and the inhibitors on a microarray in successive spotting steps using a contact robotic printer.[208]



Scheme 3.5: Fluorogenic ACC peptide microarray

Another method based on the use of PNA-encoded fluorogenic substrates that combines the advantages of solution-phase enzymatic assays and microarray technology has been reported.[157] A split and mix combinatorial library of 192 rhodamine tetrapeptide substrates encoded by 14-mer PNA tags was synthesised by alternating the use of Fmoc-protected PNA monomers and Aloc-protected amino acids.[202] Rhodamine was used as a latent fluorophore and an increase in fluorescence was only observed upon cleavage of the amide bond between the substrate and rhodamine (thus the amino acid adjacent to rhodamine corresponded to the P_1 position) (Scheme 3.6). The mixture of compounds was incubated in solution with the target protease and immobilised onto a DNA microarray for deconvolution and profiling of the "non-prime side" substrate specificity of thrombin, plasmin, caspase-3 but also complex mixtures such as crude cell lysates and human blood proteases.

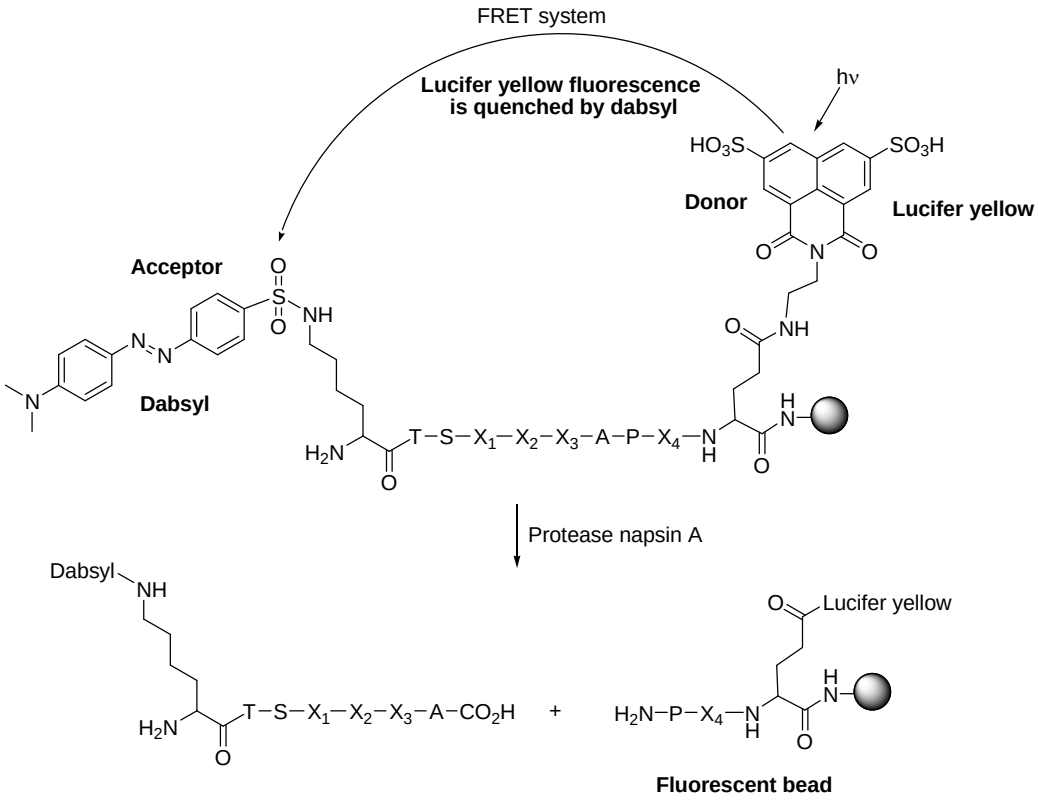


Scheme 3.6: Protease profiling using a PNA-encoded rhodamine-substrate library

3.3.6 FRET-based peptide libraries

FRET-based peptide libraries have successfully been used to study the "prime" and "non-prime side" protease substrate specificities.[209] In this strategy, a peptide is labelled with a donor fluorophore at one end and an acceptor at the other end so that the donor fluorescence is quenched by the acceptor. Following proteolytic cleavage, the donor and the acceptor are spatially separated, thus resulting in a significant increase in the donor fluorescence, which is proportional to the substrate cleavage rate. Direct proteolytic assays have been carried out on solid support-bound FRET-based peptide libraries to profile the serine protease subtilisin Carlsberg,[210] *Escherichia coli* leader peptidase and the aspartyl protease napsin A.[211] In this last example, an amino acid labelled with the donor fluorophore *N*-(2-aminoethyl)-4-amino-3,6-disulfo-1,8-naphthalimide (lucifer yellow) was first coupled to the solid support followed by the split and mix synthesis of a peptide library and final coupling with the acceptor 4-(4-dimethylaminophenylazo)benzoic acid (dabsyl). Following treatment of the library with the target protease, the most fluorescent resin beads (that corresponded to the best peptides recognised by the protease) were

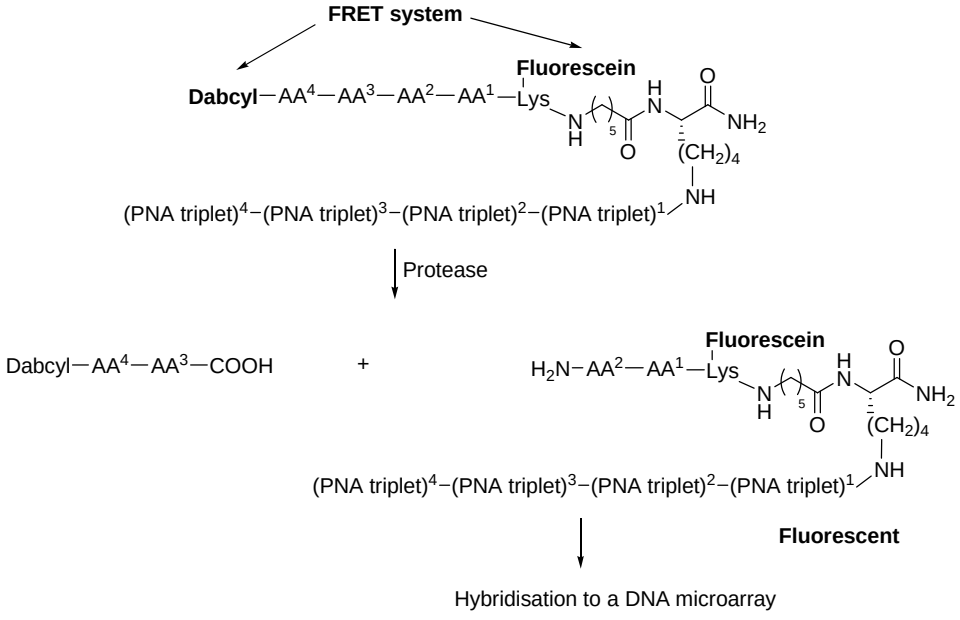
isolated and analysed by Edman degradation[133] (Scheme 3.7).[211] Petrassi investigated the "prime" and "non-prime" substrate specificities of serine, threonine and cysteine proteases.[212] A tetrapeptide ACC-positional scanning library was first used to study the specificity of the "non-prime side", which allowed the identification of optimal residues at the P_1, P_2, P_3 and P_4 positions. A FRET-based positional scanning library was subsequently prepared by coupling the identified optimal "non-prime" sequence that was labelled with a donor molecule to a previously synthesised positional scanning library containing a quencher and designed to probe the P_1', P_2', P_3' and P_4' positions.



Scheme 3.7: Use of a FRET-based peptide library to profile the protease napsin A

Recently, the development of PNA-encoded FRET-based peptide libraries have allowed for the rapid and efficient profiling of several proteases.[81, 82] A 125-member PNA-encoded peptide library was used to determine the substrate specificity of pepsin (with AA^4 = Leu) using the fluorescein (donor fluorophore) and dabcyl (internal quencher) FRET system (Scheme 3.8).[81] Each amino acid was encoded by a

PNA triplet, allowing the enzymatic assay to be carried out in solution and all members of the library (cleaved and non-cleaved peptides) to be analysed following immobilisation onto a DNA microarray. Fluorescence intensities were measured before and after proteolytic cleavage, with an increase in fluorescein intensity corresponding to a peptide cleaved by the protease (Scheme 3.8). The power of this concept was then increased *via* the introduction of internal controls (see section 2.3.3) in a 10,000-member PNA-encoded library of FRET-based peptides (fluorescein was used as the donor fluorophore and TAMRA as the quencher and internal control) (Figure 1.13).[82] Using this method, two proteases (chymopapain and subtilisin) were profiled using minimal consumption of protease (60 pmole) and library (3.5 nmole). Cleavage with a protease resulted in an increase of the fluorescein fluorescence and the highest ratios of fluorescein/TAMRA corresponded to the best peptides recognised by the enzyme. A recent microarray-based approach reported that short PNA sequences can form stable PNA/DNA duplexes and a 7-mer PNA-encoded substrate labelled with the two fluorescent dyes TAMRA and 5(6)-carboxy-X-rhodamine (ROX) was used to detect the activity of the aspartic protease β-secretase involved in Alzheimer's disease, with a sensitivity ten times superior to solution-phase assays.[213]



Scheme 3.8: Protease profiling using a PNA-encoded FRET-based peptide library

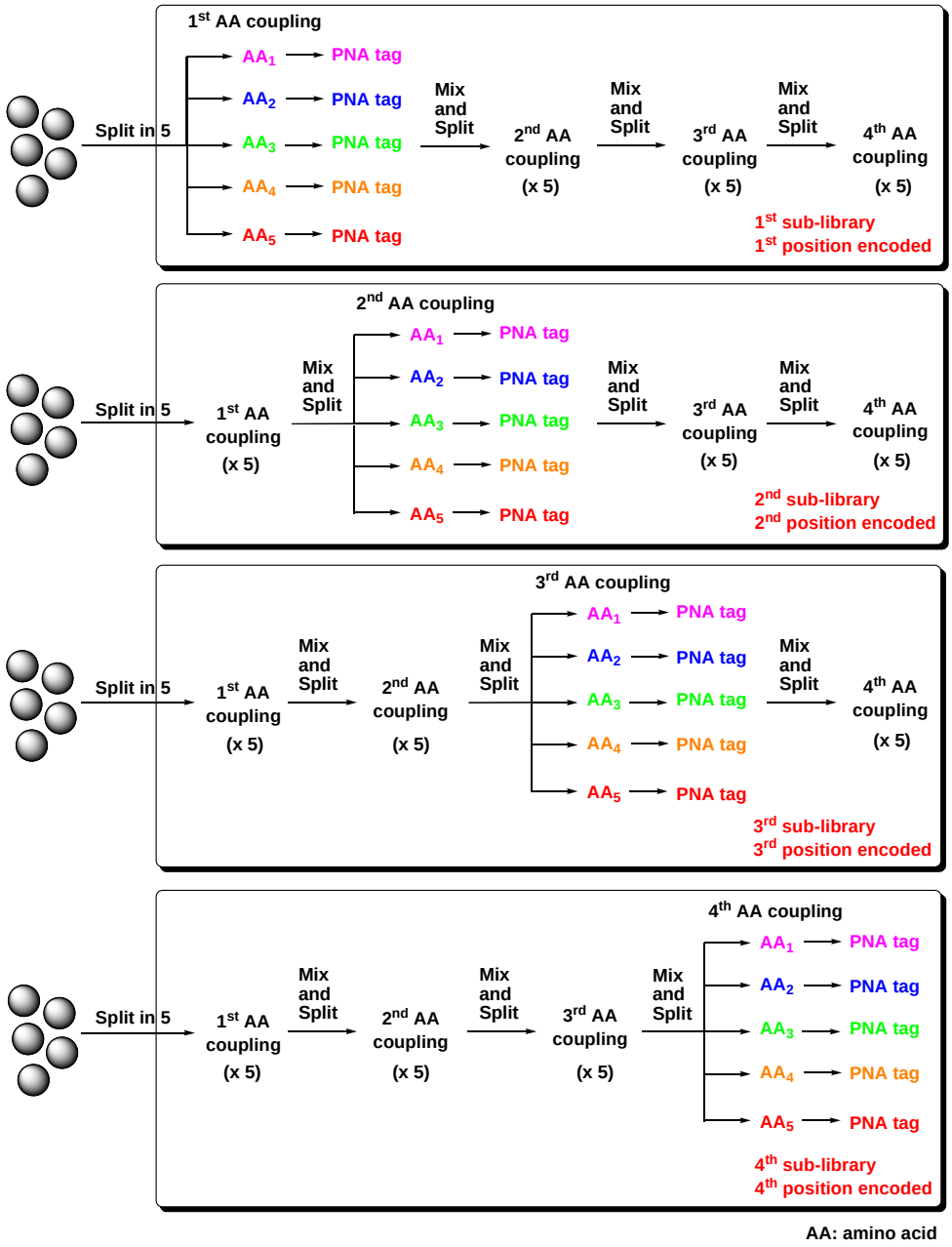
3.4 PNA-encoded FRET-based positional scanning libraries

3.4.1 Aim of the project

PNA-encoded FRET-based peptide libraries have been successfully used to determine the substrate specificity of not only kinases but also proteases.[81, 82] The aim of the project was to build a PNA-encoded positional scanning library of FRET-based tetrapeptides to carry out microarray-based proteolytic profiling. This new concept would combine the advantages of different approaches previously reported by Houghten (positional scanning libraries[61]), Ellman (positional scanning libraries of fluorogenic substrates[204]) and Bradley (PNA-encoded FRET-based peptide libraries[81, 82]). Thanks to the use of PNA tags and the microarray format, the enzymatic assays would be performed in solution using minimal quantities of enzyme and library before immobilisation of the PNA-encoded substrates to a DNA microarray for deconvolution and identification of the cleaved peptides. The use of a FRET-based library would allow the study of the "prime" and "non-prime side" protease substrate specificity. Similarly to the 10,000-member PNA-encoded FRET-based peptide library reported by Bradley,[82] a dual colour approach would be used to introduce a FRET system and internal controls within the library. By synthesising a PNA-encoded positional scanning library, the preferred amino acids for each position would be identified and only the position of the fixed amino acid would be encoded with a PNA tag so that the number of synthetic steps and PNA tags required for the synthesis of very large libraries would be significantly reduced. For example, a 10,000-member PNA-encoded combinatorial tetrapeptide library requires the use of 10,000 PNA tags to encode each peptide, whereas only 40 PNA tags are necessary in a positional scanning library, thus avoiding the use of expensive customised DNA microarrays and facilitating microarray data analysis. This chapter reports the syntheses and the enzymatic assays carried out using a 625-member PNA-encoded positional scanning library (first sub-library) and a 50,625-member PNA-encoded positional scanning library.

3.4.2 General concept of PNA-encoded positional scanning libraries

The general concept of PNA-encoded positional scanning libraries is illustrated using the example of a 625-member PNA-encoded tetrapeptide positional scanning library containing five different amino acids at each of the four positions in the library. The library is divided into four separate sub-libraries, where in each sub-library one position within the tetrapeptide is fixed with one of the five amino acids and encoded by a PNA tag, while the other three positions contain an equimolar mixture of the same five amino acids. The library is built on solid phase using a split and mix methodology. For the first sub-library synthesis, after splitting the resin into five, each batch is coupled with one of the five amino acids used as building blocks and encoded by five different PNA tags. Each pool is then mixed and split again into five. Following coupling of three additional amino acids using traditional split and mix approaches, a pot containing 625 different tetrapeptides is encoded by five PNA tags. The second, third and fourth sub-libraries are synthesised in a similar manner except that amino acids in the second, third and fourth positions are encoded by five different PNA tags. It is noteworthy that all the PNA tags are different from each other since the final objective is to hybridise the whole library in a straightforward manner on a single DNA microarray to profile protease substrate specificity and identify the preferred amino acids at a specific position. Moreover, each sub-library is composed of the same number of compounds that differ only from the position of the encoded amino acids. Thus, a PNA-encoded positional scanning library containing 625 different tetrapeptides would be encoded by 20 PNA tags (Scheme 3.9).

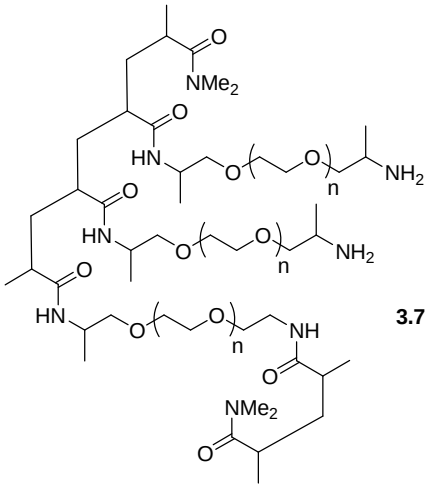


Scheme 3.9: Schematic representation of a 625-member PNA-encoded tetrapeptide positional scanning library

3.4.3 Synthetic strategy

Polyethylene glycol dimethyl acrylamide (PEGA) resin **3.7**, developed in 1992 by Meldal,[214] was used to construct the libraries. PEGA resin is obtained by copolymerisation of bis-2-acrylamidoprop-1-yl-PEG, 2-acrylamidoprop-1-yl[2-aminoprop-1-yl]-PEG and dimethyl acrylamide (Figure 3.3). PEGA resin was chosen due to its excellent swelling properties in both organic (DMF, DCM, TFA, alcohols)

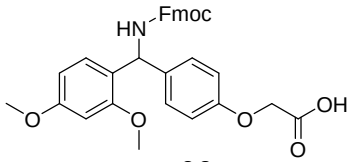
and aqueous media. It is also a highly flexible (presence of a high content of PEG), biocompatible and polar resin that favours the solubilisation of peptides and the introduction of polar compounds inside the bead.



3.7

Figure 3.3: Chemical structure of $PEGA_{1900}$ resin (n = 43)

To enable final cleavage of the products, an acid labile linker such as the Fmoc-Rink amide linker[215] **3.8** was necessary (Figure 3.4). The next important criterion in the design of the libraries was to choose specific protecting groups to carry out PNA and amino acid couplings under orthogonal conditions. Since global acidic cleavage conditions were selected, and final cleavage conditions should remove the compounds from the solid support and cleave all the protecting groups, it was necessary to choose base-labile protecting groups. Amino acids were protected with an Fmoc group since Fmoc-protected amino acids are commercially available.



3.8

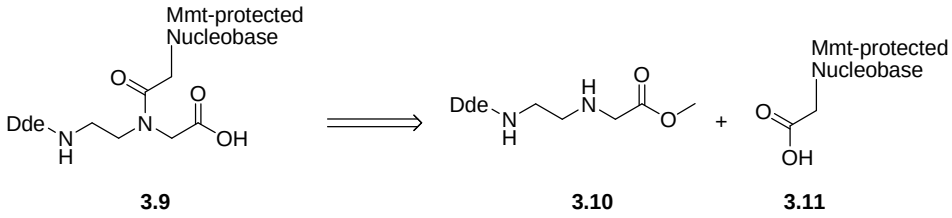
Figure 3.4: Chemical structure of the Fmoc-Rink-amide linker

For PNA synthesis, PNA monomers were used, where the nucleobases (except thymine) were protected with the acid-labile protecting group Mmt and the amino group of the backbone with Dde,[4] fully orthogonal to the Fmoc group (see section 1.1.3).[20] It was shown that the fastest cleavage kinetics in terms of Dde deprotection was obtained using PEGA resin (complete deprotection in one hour). To introduce a FRET system in the library in a similar way as the technique described for the 10,000-member PNA-encoded FRET-based peptide library synthesis,[82] one more orthogonal protecting group was required to protect the amino side chain of lysine during the positional scanning library synthesis and later on deprotect it to couple the donor fluorophore. It had to be orthogonal to basic and acidic conditions so that the palladium-labile protecting group Aloc was chosen.

3.5 Synthesis of Dde/Mmt-protected PNA monomers

3.5.1 Retrosynthesis of Dde/Mmt-protected PNA monomers

The four Dde/Mmt-protected PNA monomers **3.9** were synthesised by coupling the Dde-protected backbone **3.10** to the corresponding Mmt-protected derivatised nucleobase **3.11**, with subsequent saponification of the Dde-protected backbone's methyl ester (Scheme 3.10)**.**[4]



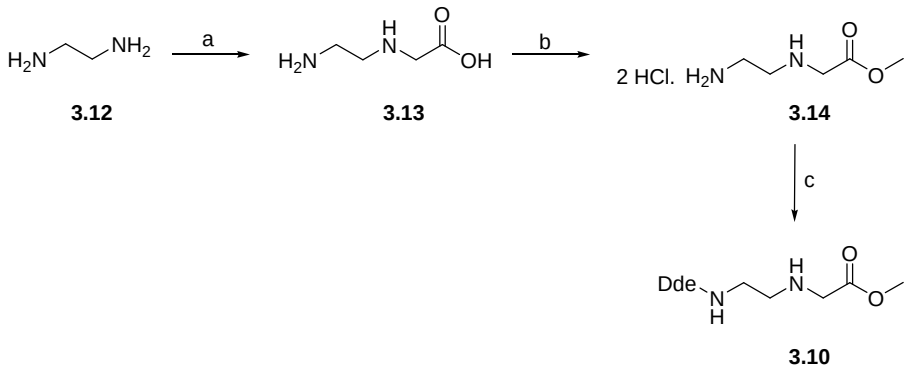
3.9 **3.10** **3.11**

Scheme 3.10: Retrosynthesis of Dde/Mmt-protected PNA monomers (No Mmt amino protection is required for thymine)

3.5.2 Synthesis of the Dde-protected backbone

Condensation of 1,2-ethanediamine **3.12** with chloroacetic acid afforded **3.13** in 66% yield. Compound **3.13** was esterified with thionyl chloride in methanol to give **3.14** in 88% yield. The primary amine of **3.14** was selectively protected with a Dde group, using Dde-OH **3.15** (previously synthesised from acylation of dimedone with

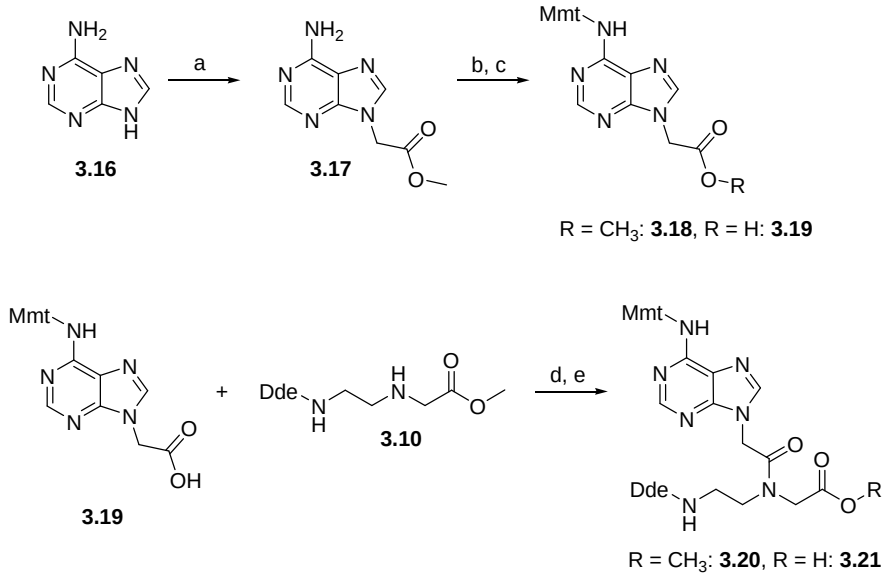
DCC/DMAP activated acetic acid)[216] in the presence of DIPEA to afford the Dde-protected backbone **3.10** in 64% yield (Scheme 3.11).[4]



Scheme 3.11: Synthesis of the Dde-protected backbone. Reagents and conditions: (a) $ClCH_2CO_2H$ (0.1 eq), 0 °C then 25 °C, 16 h; (b) $SOCl_2$ (5 eq), MeOH, 0 °C then reflux, 16 h; (c) Dde-OH **3.15** (1 eq), DIPEA (2 eq), DCM, 16 h

3.5.3 Synthesis of the adenine PNA monomer

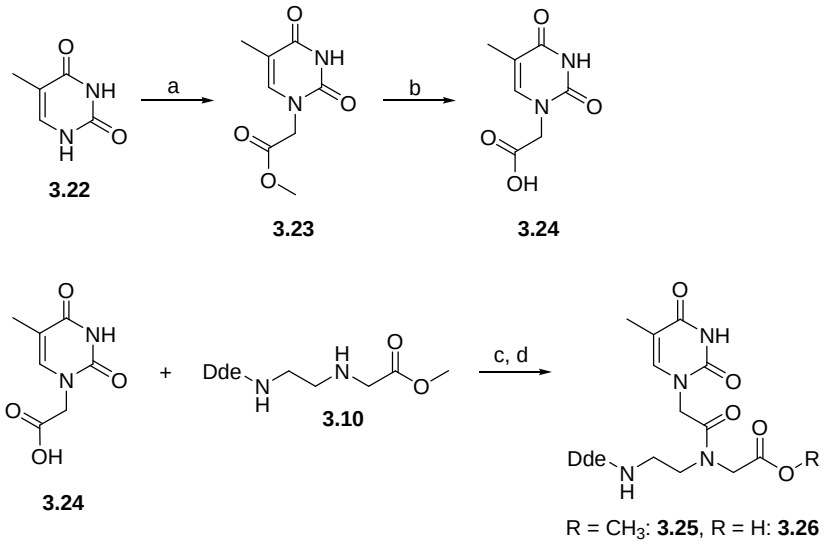
The exocyclic amino group of adenine **3.16** had to be protected to avoid any side reactions occurring at this position. Moreover, to obtain selective N^9 over N^7 alkylation, adenine **3.16** was first alkylated with methyl 2-bromoacetate, in the presence of NaH to afford **3.17** in 50% yield.[14] Protection of the primary amine with Mmt was carried out using Mmt chloride in the presence of *N*-ethylmorpholine (NEM) in pyridine/DCM followed by purification (removal of the by-product Mmt-OH) to give **3.18** in 82% yield. Saponification of **3.18** with NaOH afforded in 86% yield the derivatised adenine **3.19**, which was coupled to the Dde-protected backbone **3.10** using HBTU as a coupling reagent to give **3.20** in 85% yield. This coupling step was also carried out using PyBOP as the coupling agent, but HBTU allowed for better yields and shorter reaction times. A final saponification afforded the adenine PNA monomer **3.21** in 72% yield (Scheme 3.12). It was necessary to carry out several recrystallisations to obtain a 100% pure compound. Highly pure PNA monomers were required since the PNA-encoded libraries were synthesised on solid phase using a split and mix methodology, generating mixtures of compounds that could not be individually purified.



Scheme 3.12: Synthesis of the adenine PNA monomer. Reagents and conditions: (a) $BrCH_2CO_2CH_3$ (2 eq), NaH (1.1 eq), DMF, 25 °C, 16 h; (b) Mmt-Cl (1.5 eq), NEM (1 eq), pyridine/DCM (1/1, v/v), 40 °C, 16 h; (c) 1 M aqueous NaOH, reflux 2 h then 25 °C, 2 h; (d) HBTU (1.2 eq), DIPEA (1 eq), CH_3CN, 25 °C, 1 h; (e) MeOH/2 M aqueous Cs_2CO_3 (2/1, v/v), 25 °C, 1 h

3.5.4 Synthesis of the thymine PNA monomer

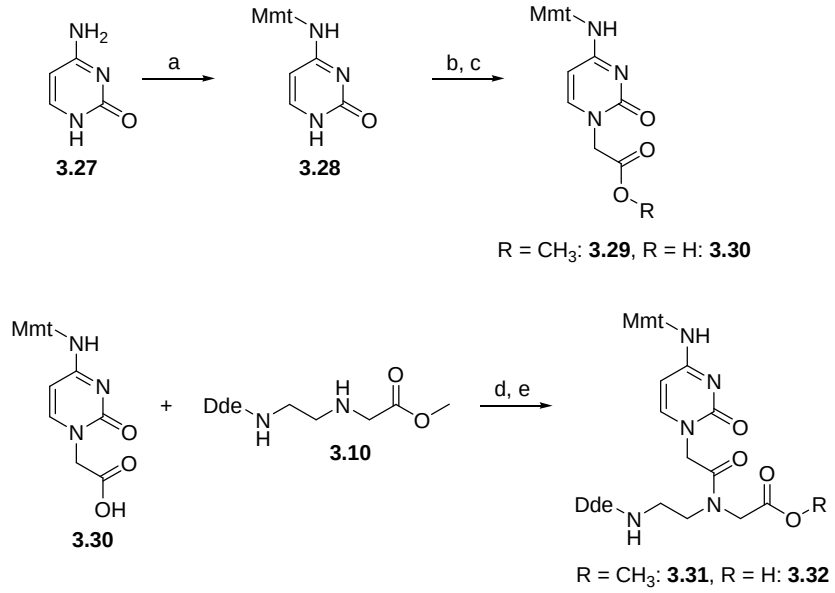
Thymine **3.22** was first alkylated with methyl 2-bromoacetate, in the presence of K_2CO_3 to afford **3.23** in 58% yield, followed by saponification with NaOH to afford the derivatised thymine **3.24** in 50% yield.[14] Similarly to the adenine PNA monomer **3.21**, the derivatised thymine **3.24** was coupled to the Dde-protected backbone **3.10** to give **3.25** in 77% yield, followed by saponification to afford the thymine PNA monomer **3.26** in 40% yield.[4] The low yields are explained by the poor solubility of thymine derivatives in organic solvents (Scheme 3.13).



Scheme 3.13: Synthesis of the thymine PNA monomer. Reagents and conditions: (a) $BrCH_2CO_2CH_3$ (1.3 eq), K_2CO_3 (1 eq), DMF, 25 °C, 16 h; (b) 2 M aqueous NaOH, reflux, 3 h; (c) HBTU (1.2 eq), DIPEA (1 eq), CH_3CN, 25 °C, 1 h; (d) MeOH/2 M aqueous Cs_2CO_3 (1/1, v/v), 25 °C, 1 h

3.5.5 Synthesis of the cytosine PNA monomer

To improve cytosine's solubility in organic solvents, the exocyclic amino group of cytosine **3.27** was first protected using Mmt chloride in the presence of NEM in pyridine/DCM yielding **3.28** in 84% yield,[19] followed by alkylation of **3.28** with methyl 2-bromoacetate and NaH to afford **3.29** in 87% yield. In this case, the by-product Mmt-OH was easily removed by recrystallisations in diethyl ether. Saponification of **3.29** gave the derivatised cytosine nucleobase **3.30** in 74% yield. In this case, PyBOP was used as the coupling reagent to couple **3.30** to the backbone **3.10** giving **3.31** in 57% yield. HBTU was not used as it was more difficult to purify the product and lower yields were obtained. Compound **3.31** was saponified and purified in a similar way to the adenine PNA monomer **3.21** to afford the cytosine PNA monomer **3.32** in 77% yield (Scheme 3.14).[4]



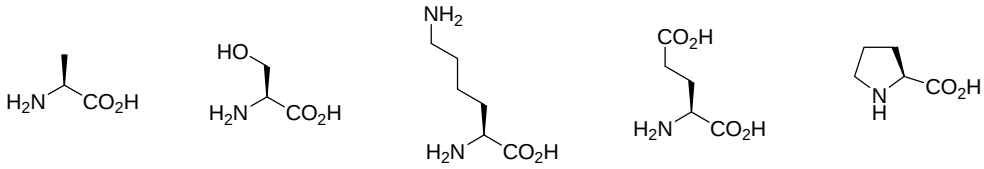
Scheme 3.14: Synthesis of the cytosine PNA monomer. Reagents and conditions: (a) Mmt-Cl (1.5 eq), NEM (1 eq), pyridine/DCM (1/1, v/v), 50 °C, 16 h; (b) $BrCH_2CO_2CH_3$ (2 eq), NaH (1.1 eq), DMF, 25 °C, 16 h; (c) 2 M aqueous NaOH, reflux 2 h; (d) PyBOP (1.1 eq), DIPEA (2 eq), DMF, 25 °C, 16 h; (e) MeOH/2 M aqueous Cs_2CO_3 (2/1, v/v), 25 °C, 1 h

3.6 Synthesis of a 625-member PNA-encoded positional scanning library

As proof-of-principle, the synthesis of the first sub-library of a 625-member PNA-encoded tetrapeptide positional scanning library (PNA encoding of five amino acids located in the first position of the tetrapeptide) was first investigated.

3.6.1 Selection of amino acids

To introduce good "diversity" into the library, five natural amino acids with differing properties were chosen: alanine **3.33** (hydrophobic), serine **3.34** (hydrophilic), lysine **3.35** (basic and hydrophilic), glutamic acid **3.36** (acidic and hydrophilic) and proline **3.37** (hydrophobic and cyclic) (Figure 3.5). The side chains of serine **3.34** and glutamic acid **3.36** were protected with tert-butyl groups, while a Boc group was chosen for lysine **3.35**.



Alanine **3.33** Serine **3.34** Lysine **3.35** Glutamic acid **3.36** Proline **3.37**

Figure 3.5: Chemical structures of the five selected amino acids

3.6.2 Design of the PNA tags

A 12-mer PNA has been proven to be a suitable length for hybridisation, and the formation of strong PNA/DNA duplexes with good mismatch discrimination.[77, 81] PNA tags were composed of twelve PNA monomers, with the first seven PNA monomers common to all the amino acids encoded within the library, and a quintuplet PNA used as the encoding sequence. PNA tags were designed with specific criteria.[217] To avoid aggregation and solubility issues in aqueous solutions, the purine content was limited to 60% and sequences did not contain more than four consecutive adenines or three contiguous guanines. Moreover, it was essential not to design self-complementary PNA sequences with inverse repeats, hairpins or palindromes since it leads to aggregation with formation of very stable PNA/PNA duplexes. The presence of a common 7-mer PNA was important to have PNA tags with similar T_m and obtain homogeneous detection of the library members after hybridisation. The PNA tags differed by at least three base pair mismatches to ensure effective and selective hybridisation, where every PNA tag hybridised with its complementary DNA immobilised on a DNA microarray. Following these basic rules, a common 7-mer PNA with the sequence GATCATC (from N → C) was designed as well as the 5-mer PNA tags used to encode the five amino acids in the first position of the tetrapeptide (Table 3.1). The guanine PNA monomer used to prepare these PNA tags was synthesised by Diaz-Mochon.[4]

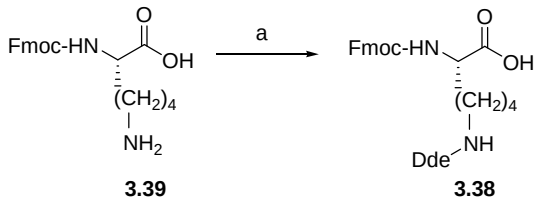
Table 3.1: Sequences of the 5-mer PNA tags of the first sub-library

Encoded amino acid	5-mer PNA tag (from N → C)
Ala	GACTT
Ser	CGGTT
Lys	AACTC
Glu	TCATC
Pro	CTGTC

3.6.3 Synthesis of the common part of the library

3.6.3.1 Synthesis of Fmoc-Lys(Dde)-OH

The common part of the library was synthesised using Fmoc-Lys(Dde)-OH **3.38** as a core. This core allowed for attachment to the solid support through the carboxylic acid functionality, the construction of the PNA tag using Dde chemistry[20] on the side chain amine and Fmoc-based SPPS through the α-amino group. The condensation of Fmoc-Lys-OH **3.39** with Dde-OH **3.15** in refluxing ethanol and TFA for 60 h afforded Fmoc-Lys(Dde)-OH **3.38** in 69% yield (Scheme 3.15).[218]

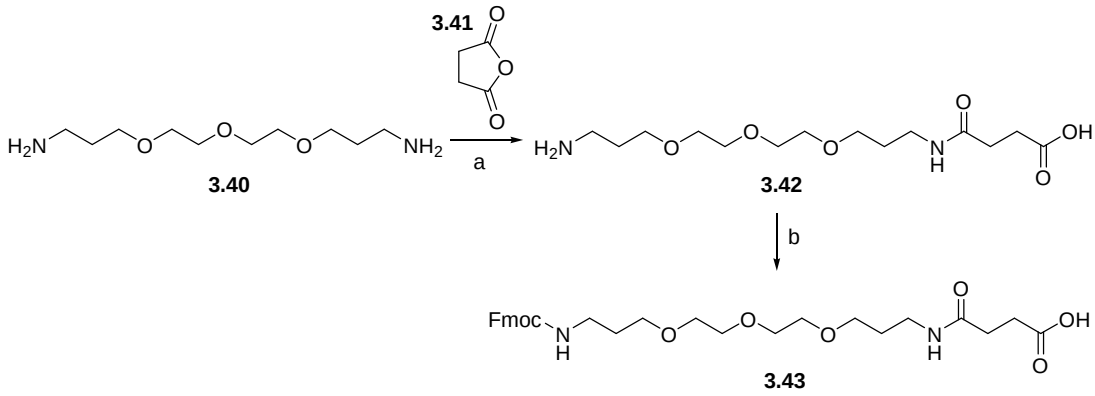


3.39 **3.38**

Scheme 3.15: Synthesis of Fmoc-Lys(Dde)-OH. Reagents and conditions: (a) Dde-OH **3.15** (2 eq), TFA (0.1 eq), EtOH, reflux, 60 h

3.6.3.2 Synthesis of Fmoc-PEG spacer-OH

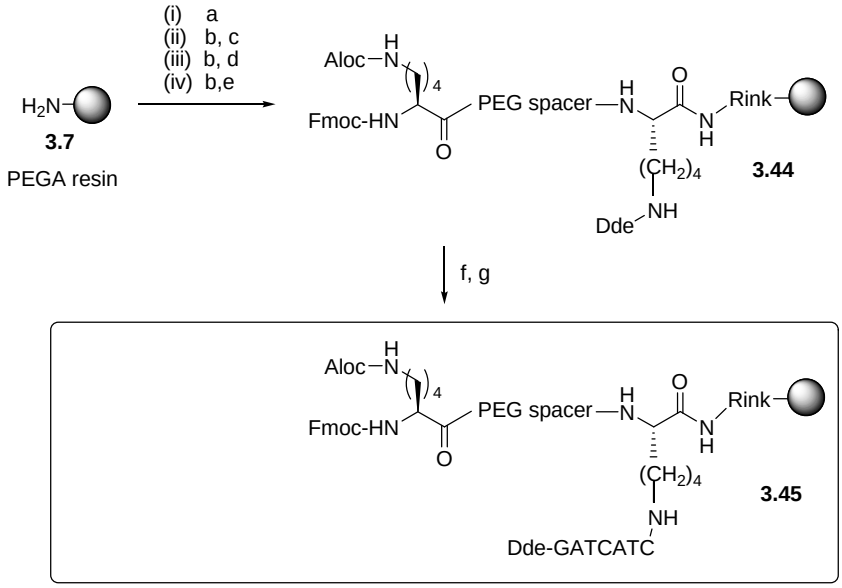
A hydrophilic and highly flexible spacer such as PEG was also connected to the resin as it increased the solubility of every library member. Moreover, if the peptides and their PNA tags were too close in proximity, undesirable interactions could appear and the cleavage site might not be fully accessible to the protease. Nucleophilic attack by one primary amine of the starting compound **3.40** on succinic anhydride **3.41** gave **3.42**. Following protection of the remaining free amine with an Fmoc group, Fmoc-PEG spacer-OH **3.43** was afforded in 74% yield (Scheme 3.16).[219]



Scheme 3.16: Synthesis of Fmoc-PEG spacer-OH. Reagents and conditions: (a) CH_3CN, 25 °C, 3 h; (b) Fmoc-OSu (1.3 eq), CH_3CN, DIPEA, pH 8-9, 25 °C, 16 h

3.6.3.3 Synthesis of the common part of the library

In the first step, the Fmoc-Rink amide linker **3.8** was loaded onto PEGA resin **3.7**, which was followed by removal of the Fmoc group and coupling with Fmoc-Lys(Dde)-OH **3.38**. In the last two steps, the Fmoc group was first deprotected followed by stepwise couplings with the Fmoc-PEG spacer-OH **3.43** and the commercially available Fmoc-Lys(Aloc)-OH to afford compound **3.44**. The construction of the 7-mer PNA with the sequence GATCATC (from N → C) was carried out on the amino side chain of Fmoc-Lys(Dde)-OH **3.38** using Dde chemistry.[20] The common part of the library **3.45** was thus obtained. All the couplings were carried out using PyBOP as the coupling reagent and completion of each reaction was verified by running a ninhydrin test[220] (Scheme 3.17).



Scheme 3.17: Synthesis of the common part of the library. Reagents and conditions: (a) Fmoc-Rink-amide linker **3.8** (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; (b) 20% piperidine in DMF, 2 cycles of 10 min, 25 °C; (c) Fmoc-Lys(Dde)-OH **3.38** (5 eq), PyBOP (4.8 eq)/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq), 25 °C, 6 h; (d) Fmoc-PEG spacer-OH **3.43** (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; (e) Fmoc-Lys(Aloc)-OH (5 eq), PyBOP (4.8 eq)/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq), 25 °C, 6 h; (f) $NH_2OH.HCl$ (1.3 eq)/imidazole (1 eq), NMP/DMF, 25 °C, 1 h; (g) Dde/Mmt protected PNA (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; Repeat 6 times steps f and g

An analytical amount of the common part of the library **3.45** was cleaved from the resin using a mixture of TFA/TIS/DCM (90/5/5). The HPLC trace of the crude compound showed an excellent purity (91%), which implied over 99% purity per chemical step (22 steps including deprotection steps). MALDI-TOF MS analysis showed the mass of the expected compound (Figure 3.6). The purity and the small number of deletion sequences confirmed the efficiency of the PNA coupling steps.[4]

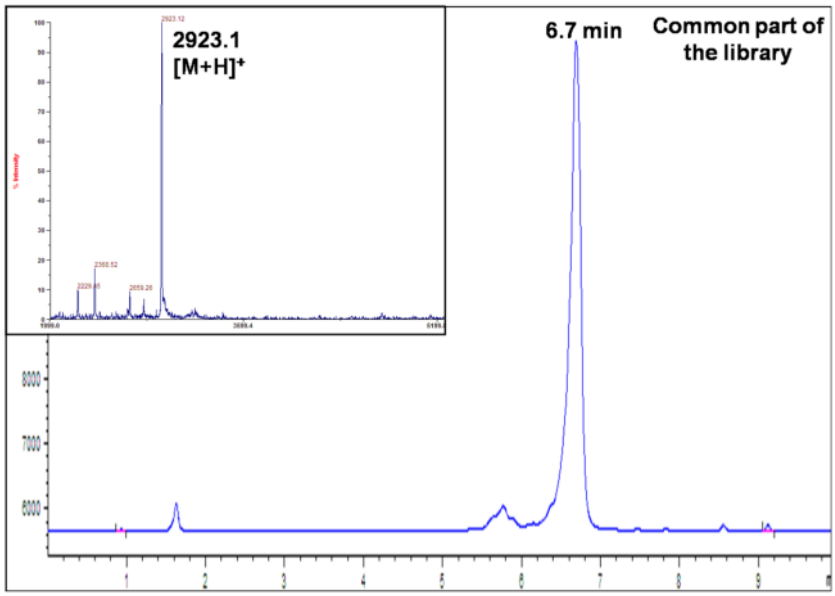
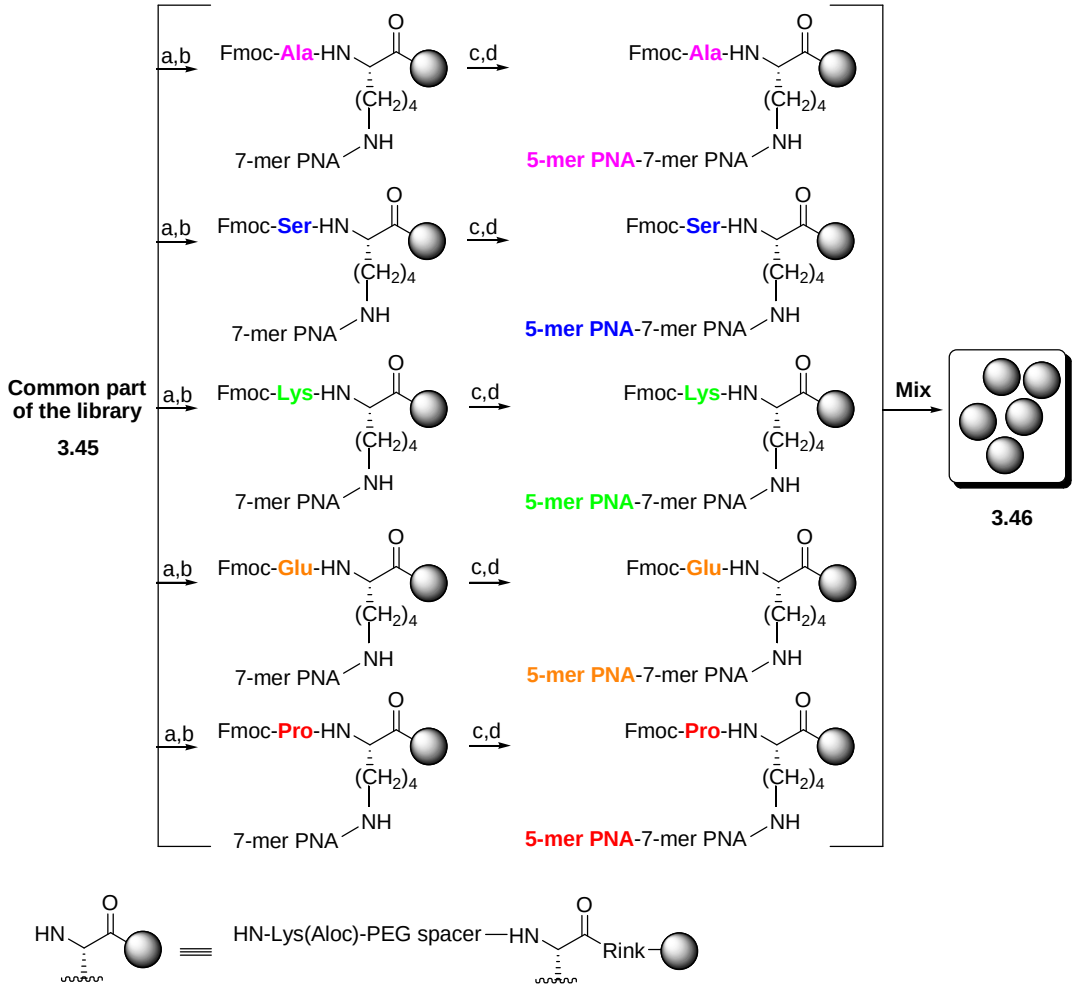


Figure 3.6: HPLC spectrum (ELSD, 91% purity) and MALDI-TOF MS spectrum, top left of the common part of the library **3.45** after acidic cleavage

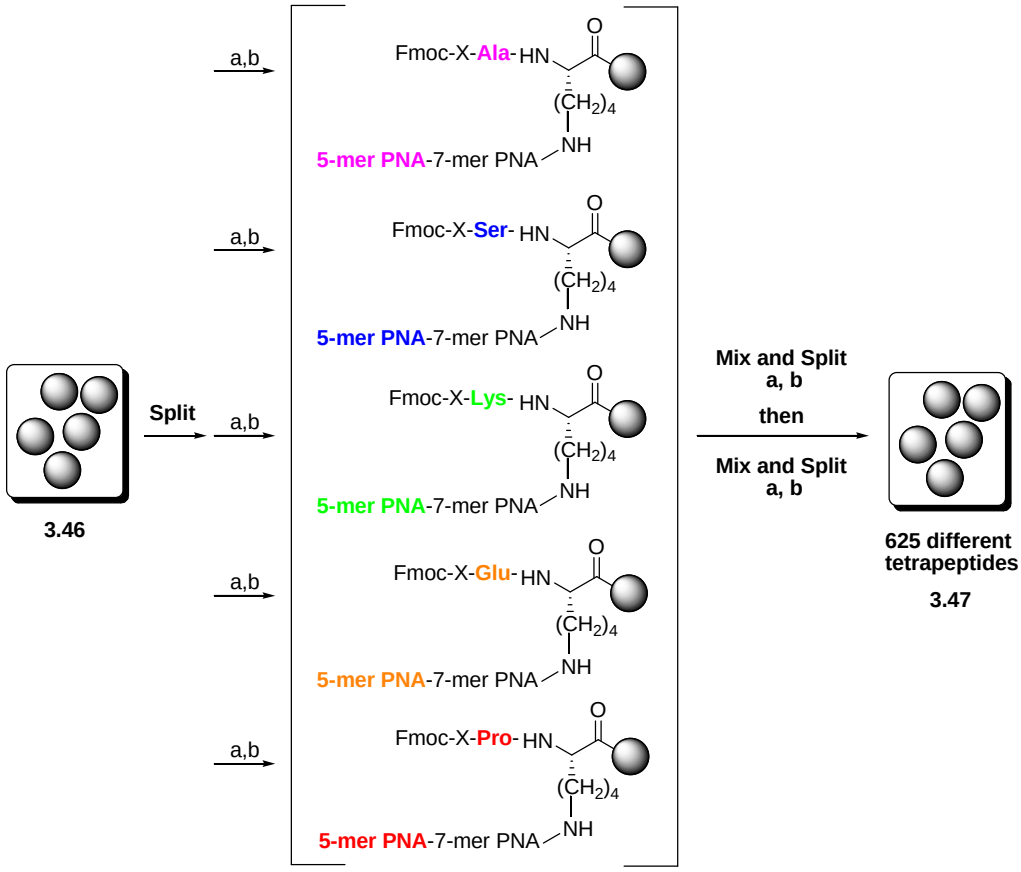
3.6.4 Synthesis of the first sub-library

The synthesis of the 625-member PNA-encoded positional scanning library (first sub-library) was carried out. In this section, all the couplings were carried out using PyBOP as the coupling reagent and completion of each reaction was verified by running a ninhydrin test.[220] The common part of the library **3.45** was split into five pools. Each batch was subjected to Fmoc deprotection followed by coupling with one of the five selected Fmoc-protected amino acids (Ala, Ser, Lys, Glu and Pro). After Dde deprotection[20] on each batch, the encoding PNA arm was built to encode the first position of one amino acid in the tetrapeptide. Each pool contained a different amino acid and was consequently encoded with a different 5-mer PNA tag (Table 3.1). After mixing all the pools, a mixture of compounds **3.46** was obtained. In the two following schematic representations 3.18 and 3.19, the seven PNA monomers are abbreviated to "7-mer PNA", while the five different PNA monomers used as encoding tags are noted "5-mer PNA". Moreover, the general structure of the library is also simplified by representing only the lateral chain of the core lysine onto which the PNA tag is built (Scheme 3.18).



Scheme 3.18: Schematic representation of the PNA encoding of the first position in the tetrapeptide. Reagents and conditions: (a) 20% piperidine in DMF, 2 cycles of 10 min, 25 °C; (b) Fmoc-AA-OH (5 eq), PyBOP (4.8 eq)/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq), 25 °C, 6 h (where AA = amino acid); (c) $NH_2OH.HCl$ (1.3 eq)/imidazole (1 eq), NMP/DMF, 25 °C, 1 h; (d) Dde/Mmt protected PNA (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; Repeat 4 times steps c and d

The resulting mixture **3.46** was then split into five different pools. Following Fmoc deprotection, each batch was coupled with one of the five Fmoc-protected amino acids and mixed again. This procedure was repeated twice to end up with a pot **3.47** containing 625 different tetrapeptides encoded by five PNA tags (Scheme 3.19). The general structure of the first sub-library **3.47** that is obtained at the end of the split and mix reactions is displayed in Figure 3.7.



Scheme 3.19: Schematic representation of the 625-member PNA-encoded positional scanning library (first sub-library), where X = equimolar mixture of five amino acids. Reagents and conditions: (a) 20% piperidine in DMF, 2 cycles of 10 min, 25 °C; (b) Fmoc-AA-OH (5 eq), PyBOP (4.8 eq)/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq), 25 °C, 6 h (where AA = amino acid)

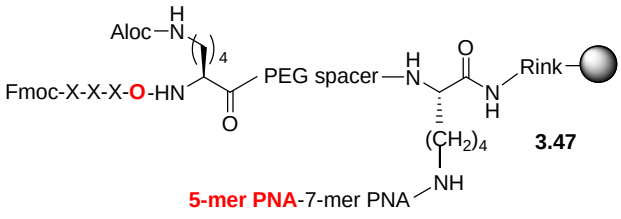
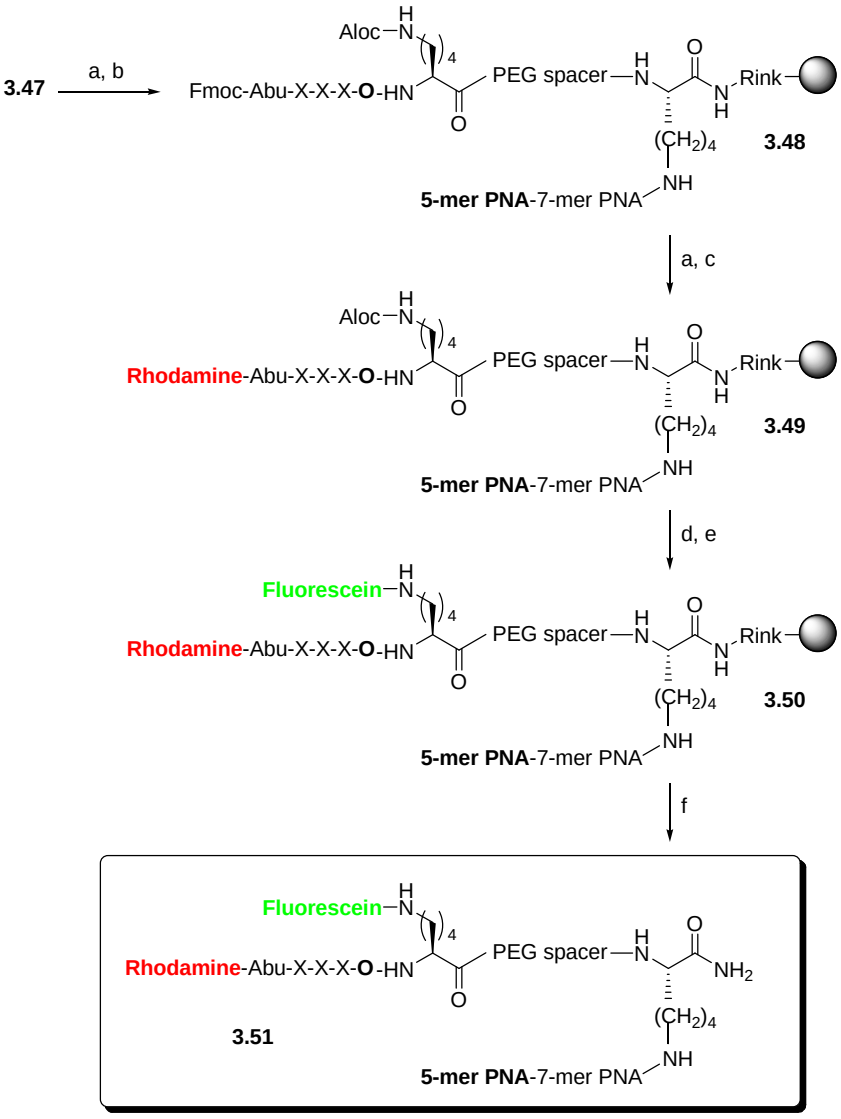


Figure 3.7: General structure of the 625-member PNA-encoded positional scanning library (first sub-library) before the introduction of the FRET system, where O = one of the five encoded amino acids and X = equimolar mixture of five amino acids

To detect and monitor proteolytic activity, a FRET system was introduced within the library to produce an internally quenched fluorescent peptide. An important criterion was to choose the correct donor fluorophore/acceptor fluorophore (quencher) couple so that the emission spectrum of the donor fluorophore overlaps with the absorption spectrum of the quencher. The donor/quencher couple 5(6)-carboxyfluorescein

(fluorescein)/rhodamine was chosen, with fluorescein being effectively quenched by the nearby rhodamine acceptor in a FRET system. Fluorescein was selected as the donor (λ_{ex} = 495 nm and λ_{em} = 520 nm) and rhodamine as the acceptor (λ_{ex} = 545 nm and λ_{em} = 580 nm) since these two dyes were biocompatible, cheap and fluorescence could be easily detected in most instruments using standard filters.

It has been previously reported that the introduction of a spacer at the end of the peptide sequence allowed better hydrolysis kinetics.[221] However, if the spacer is too long then the FRET system would not work properly as the two fluorophores would be too far away from each other. If it is too short then the protease would not have good access to the cleavage site. Consequently the 4-carbon spacer γ-amino butyric acid (γ-Abu) was used. Following Fmoc deprotection, the commercially available Fmoc-γ-Abu was coupled to resin **3.47** to give the expected compound **3.48**. Finally, the FRET system was incorporated in the library. Fluorescein was attached to the lateral chain of lysine on the *C*-terminal side of the cleavage site, while rhodamine was attached to the *N*-terminal end. Another Fmoc deprotection and coupling with rhodamine afforded the rhodamine-labelled compound **3.49**. The Aloc group was then removed under neutral conditions using a catalytic amount of $Pd(PPh_3)_4$ in the presence of $PhSiH_3$ as a scavenger in dry DCM. In order to avoid oxidation of the palladium, the deprotection was carried out under an inert atmosphere (N_2).[222] Coupling with fluorescein with subsequent washings with 20% piperidine in DMF to remove impurities[151] afforded the internally quenched fluorescent peptide **3.50**. A final acidic cleavage (TFA/TIS/DCM: 90/5/5, 2 h) released the compounds from the solid support and simultaneously removed all the acid-labile side chain protecting groups of amino acids and PNA monomers. The first sub-library **3.51** was obtained in 67% yield (> 99% yield per chemical step (46 steps including deprotection steps)). This yield was calculated using the first sub-library **3.51** average mass of 5135.5 g/mol, obtained using both the average masses of the five amino acids and the four Dde/Mmt-protected PNA monomers (Scheme 3.20). Since a mixture of 625 different compounds was obtained, it was no longer possible to determine the identity of every library member by MALDI-TOF MS analysis.



Scheme 3.20: General structure of the 625-member PNA-encoded positional scanning library (first sub-library), where O = one of the five encoded amino acids and X = equimolar mixture of five amino acids. Reagents and conditions: (a) 20% piperidine in DMF, 2 cycles of 10 min, 25 °C; (b) Fmoc-γ-Abu-OH (5 eq), PyBOP (4.8 eq)/HOBt (5 eq), DMF (0.1 M), DIPEA (10 eq), 25 °C, 6 h; (c) Rhodamine (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; (d) $Pd(PPh_3)_4$ (0.1 eq)/$PhSiH_3$ (24 eq), dry DCM, N_2, 2 cycles of 15 min; (e) 5(6)-carboxyfluorescein (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; (f) TFA/TIS/DCM (90/5/5), 2 h

To verify the presence of the FRET system within the first sub-library, two resin beads were taken up at the end of the first sub-library synthesis (just before acidic cleavage) and scanned using fluorescein and rhodamine filters. Fluorescent images of the two resin beads **3.51** are shown in Figure 3.8, confirming the correct labelling of the first sub-library with the two dyes.

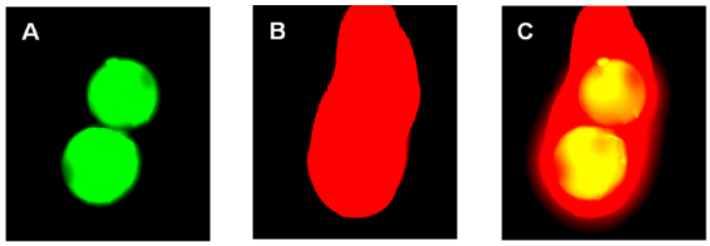
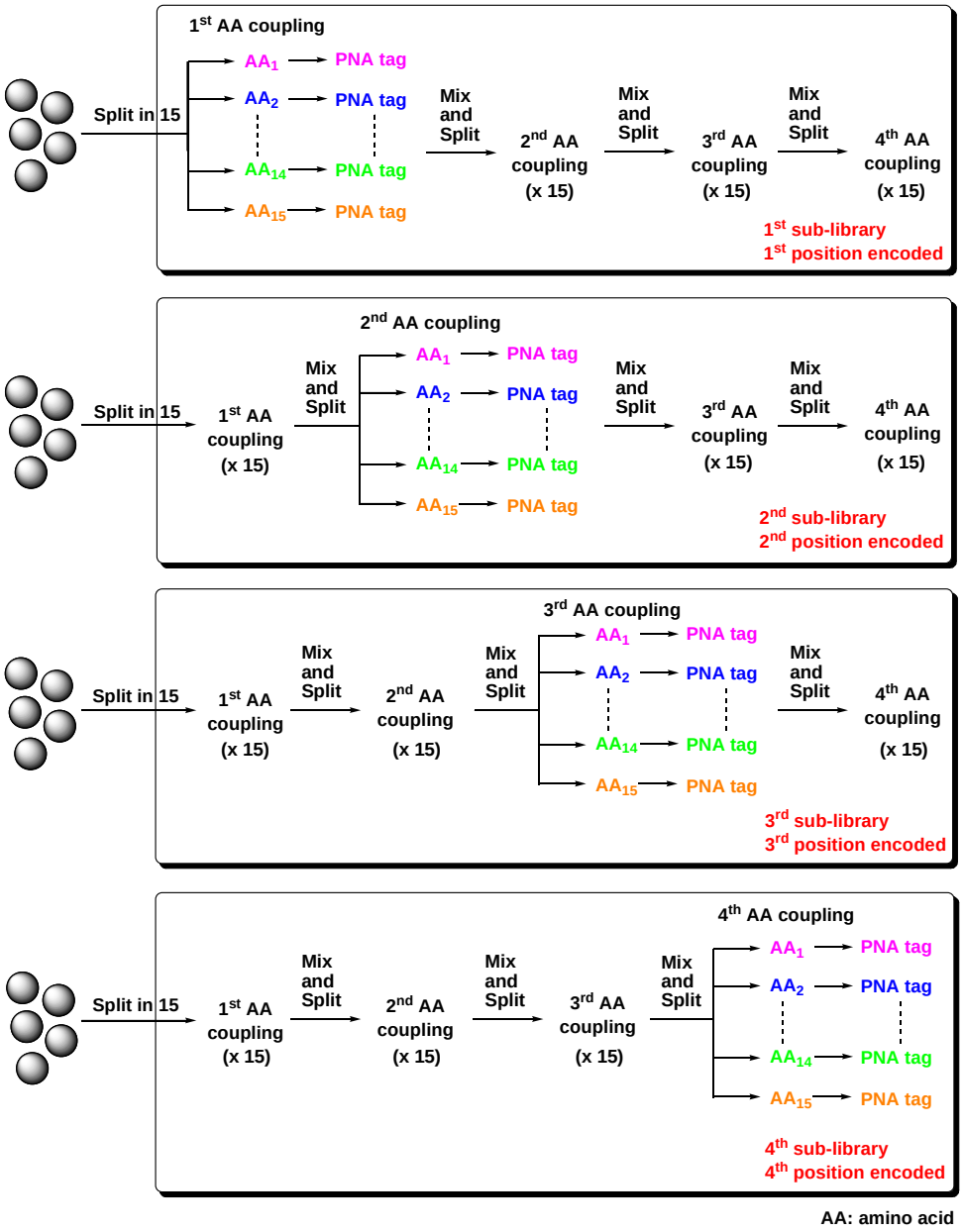


Figure 3.8: Fluorescent images of two resin beads from the 625-member PNA-encoded positional scanning library (first sub-library): (A) fluorescein image; (B) rhodamine image; (C) superimposed image of fluorescein and rhodamine images

3.7 Synthesis of a 50,625-member PNA-encoded positional scanning library

3.7.1 Introduction

The 625-member PNA-encoded positional scanning library (first sub-library) **3.51** was successfully synthesised and used to profile three different proteases (results of the enzymatic assays are reported in section 3.8.2). Consequently, the concept was extended to the synthesis of a larger PNA-encoded positional scanning library composed of 50,625 different tetrapeptides encoded for by just 60 PNA tags. To reduce the amount of PNA monomers needed to synthesise the library, fifteen amino acids were chosen from the 20 natural amino acids excluding Asn and Asp (similar side-chain structures and polarity to Gln and Glu respectively), Leu (similar side-chain structure and polarity to Ile), Cys and Met (to avoid formation of disulfide bonds and oxidation of the sulphur). The library was also divided into four separate sub-libraries, in which 15 different amino acids were held constant at one position in the tetrapeptide and encoded by a specific PNA tag, while the other three positions contained an equimolar mixture of 15 amino acids. Each sub-library contained 50,625 (15^4) different compounds encoded by 15 PNA tags, and in total only 60 different mixtures of compounds were prepared (Scheme 3.21).



Scheme 3.21: Schematic representation of a 50,625-member PNA-encoded tetrapeptide positional scanning library

3.7.2 Design of the PNA tags

To design the 60 PNA tags, the method used for the 625-member PNA-encoded positional scanning library could have been adapted to the synthesis of a larger positional scanning library by fixing a common 6-mer PNA (ATCATC, from N → C) and then encoded every amino acid with a 6-mer PNA tag. However this approach required the use of the four natural bases, and some PNA monomers such as the PNA guanine monomer are difficult to synthesise on large scale (solubility issues, poor

yields).[4] Brenner labelled each nucleic acid of a library with a DNA tag composed of only three natural bases (A, C and T) by using the following four-base "words" (TTAC, AATC, TACT, ATCA, ACAT, TCTA, CTTT, and CAAA).[223] A total number of 16,777,216 (8^8) tags of 32 bases could be used, allowing encoding of very large combinatorial libraries. It was also demonstrated that the systematic presence of three A-T and one G-C complementary base pairs between duplexes ensures the generation of isothermal T_m's. All these features appeared as an attractive strategy to design the 60 PNA tags. Each PNA tag was composed of three four-base "words" to produce a 12-mer PNA tag. To minimise PNA aggregation and avoid the preparation of PNA tags containing more than four adenines in a row, with an adenine content superior to 60%, or composed of five consecutive thymines (these PNAs are difficult to synthesise, resulting in the formation of deleted sequences[224]), the four-base "words" TTAC, AATC, CTTT and CAAA were not used. The 60 PNA tags were designed following the recommendations described in section 3.6.2. Each PNA tag was fixed with a four-base "word", different between the four sub-libraries (TCTA, ACAT, ATCA and TACT were used for the first, second, third and fourth sub-libraries respectively). Within a sub-library, a specific amino acid was thus encoded by an 8-mer PNA tag and all the PNA tags differed by at least 3 mismatches to have a good discrimination between perfect and mismatched duplexes (Table 3.2).

Table 3.2: Sequences of the PNA tags (from N → C) used to encode amino acids in the four sub-libraries

	1st sub-library	**2nd sub-library**	**3rd sub-library**	**4th sub-library**
Ala	TCTA TCTA TCTA	TCTA TCTA ACAT	TCTA TCTA ATCA	TCTA TCTA TACT
Arg	TCTA ACAT TCTA	TCTA ACAT ACAT	TCTA ACAT ATCA	TCTA ACAT TACT
Gln	TCTA ATCA TCTA	TCTA ATCA ACAT	TCTA ATCA ATCA	TCTA ATCA TACT
Glu	TCTA TACT TCTA	TCTA TACT ACAT	TCTA TACT ATCA	ACAT TCTA TACT
Gly	ACAT TCTA TCTA	ACAT TCTA ACAT	ACAT TCTA ATCA	ACAT ACAT TACT
His	ACAT ACAT TCTA	ACAT ACAT ACAT	ACAT ACAT ATCA	ACAT ATCA TACT
Ile	ACAT ATCA TCTA	ACAT ATCA ACAT	ACAT TACT ATCA	ACAT TACT TACT
Lys	ACAT TACT TCTA	ACAT TACT ACAT	ATCA TCTA ATCA	ATCA TCTA TACT
Phe	ATCA TCTA TCTA	ATCA TCTA ACAT	ATCA ACAT ATCA	ATCA ACAT TACT
Pro	ATCA ACAT TCTA	ATCA ATCA ACAT	ATCA ATCA ATCA	ATCA ATCA TACT
Ser	ATCA ATCA TCTA	ATCA TACT ACAT	ATCA TACT ATCA	ATCA TACT TACT
Thr	ATCA TACT TCTA	TACT TCTA ACAT	TACT TCTA ATCA	TACT TCTA TACT
Trp	TACT TCTA TCTA	TACT ACAT ACAT	TACT ACAT ATCA	TACT ACAT TACT
Tyr	TACT ACAT TCTA	TACT ATCA ACAT	TACT ATCA ATCA	TACT ATCA TACT
Val	TACT ATCA TCTA	TACT TACT ACAT	TACT TACT ATCA	TACT TACT TACT

3.7.3 Synthesis of the four sub-libraries

The common part of the 50,625-member PNA-encoded positional scanning library **3.44** was synthesised following the procedure described in section 3.6.3.3, and then split into four equal batches of resin, where the different four-base "words" (TCTA, ACAT, ATCA and TACT for the first, second, third and fourth sub-libraries respectively) were built on the side chain of lysine using Dde chemistry[20] to give the common part of the first **3.52**, second **3.53**, third **3.54** and fourth **3.55** sub-libraries (Figure 3.9).

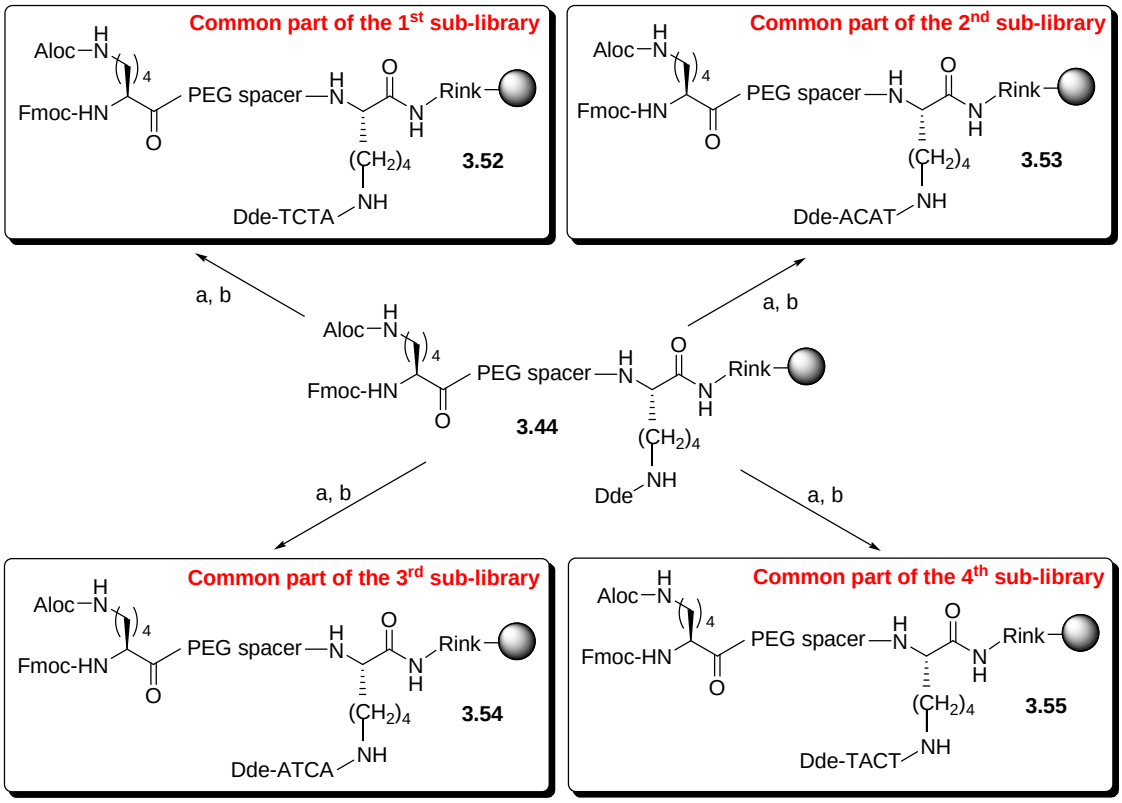


Figure 3.9: Synthesis of the common part of the four sub-libraries. Reagents and conditions: (a) $NH_2OH.HCl$ (1.3 eq)/imidazole (1 eq), NMP/DMF, 25 °C, 1 h; (b) Dde/Mmt protected PNA (5 eq), PyBOP (4.8 eq), DMF (0.1 M), NEM (10 eq), 25 °C, 6 h; Repeat 3 times steps a and b

An analytical amount of resins **3.52**, **3.53**, **3.54** and **3.55** were cleaved from the resin using a mixture of TFA/TIS/DCM (90/5/5). The HPLC traces and the MALDI-TOF MS spectra of the crude compounds are displayed in Figure 3.10. Excellent purities of 78% (> 98% purity per chemical step), 84% (99% purity per chemical step), 83% (99% purity per chemical step) and 85% (99% purity per chemical step) were found

for the first, second, third and fourth sub-libraries respectively. MALDI-TOF MS analysis showed the expected masses of the compounds. It was extremely important to carry out these analyses since the next steps of split and mix generated mixtures of compounds that could not be characterised by HPLC and MALDI-TOF MS.

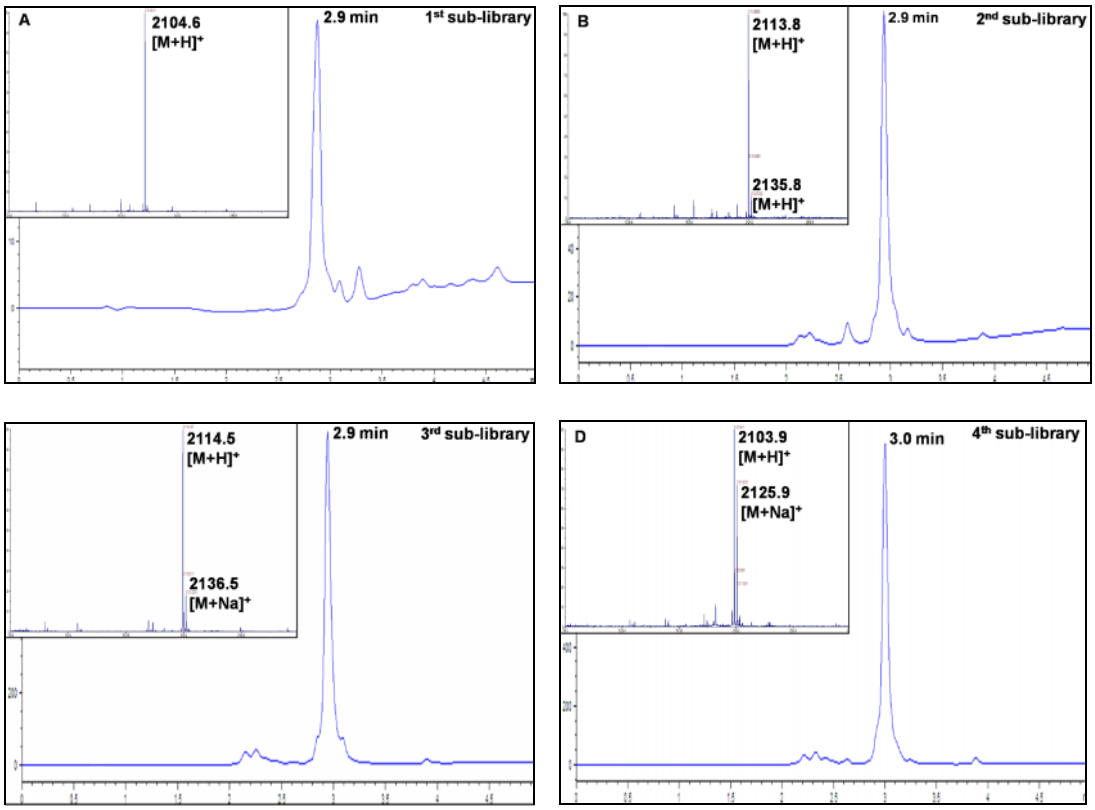


Figure 3.10: HPLC spectrum and MALDI-TOF MS spectrum, top left of the common part of (A) the first sub-library; (B) the second sub-library; (C) the third sub-library; (D) the fourth sub-library

The four sub-libraries were synthesised separately using the common part of the sub-libraries **3.52**, **3.53**, **3.54** and **3.55**, and following the procedure described for the 625-member PNA-encoded positional scanning library (see section 3.6.4), except that fifteen amino acids were coupled on each batch of resin and each amino acid was encoded by an 8-mer PNA tag. The final general structures of the four sub-libraries of the 50,625-member PNA-encoded positional scanning library **3.56**, **3.57**, **3.58** and **3.59** are displayed in Figure 3.11. Images of one resin bead of the first sub-library (obtained just before the final acidic cleavage) validated the presence of the FRET system (Figure 3.12).

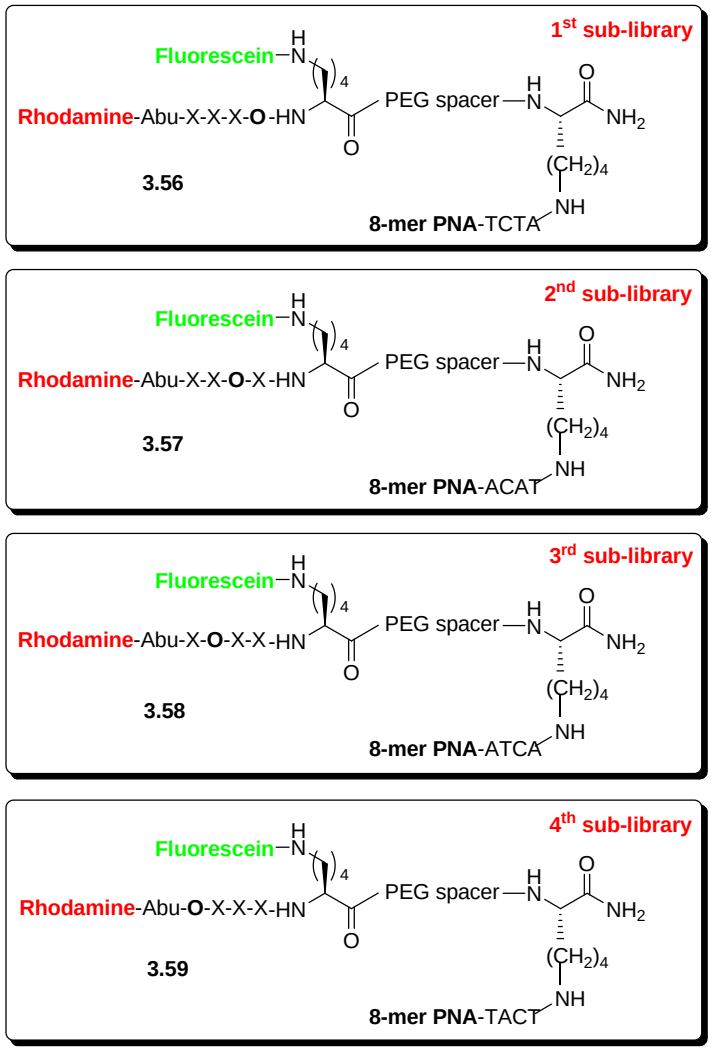


Figure 3.11: General structure of the 50,625-member PNA-encoded positional scanning library (four sub-libraries), where O = one of the fifteen encoded amino acids and X = equimolar mixture of fifteen amino acids

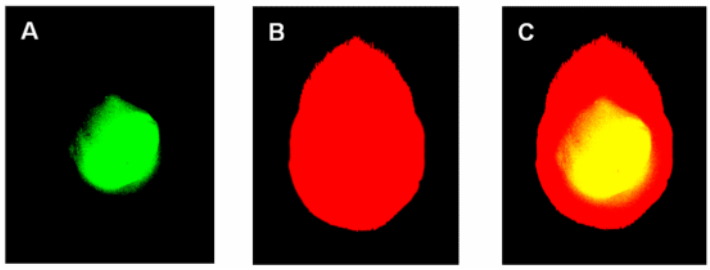


Figure 3.12: Fluorescent images of one resin bead from the 50,625-member PNA-encoded positional scanning library (first sub-library): (A) fluorescein image; (B) rhodamine image; (C) superimposed image of fluorescein and rhodamine images

3.8 Determination of protease substrate specificity

3.8.1 General concept

Following solid-phase synthesis, each peptide of the PNA-encoded positional scanning library was labelled with the donor fluorophore fluorescein at the *C*-terminus and the acceptor fluorophore rhodamine at the *N*-terminus so that the fluorescein fluorescence was quenched by rhodamine. In a typical enzymatic assay, the library was incubated in solution with the target protease (Figure 3.13) and the resulting solution was added onto a DNA microarray. Thus, both cleaved and non-cleaved peptides were hybridised onto the DNA microarray. Following washing, the microarray was scanned at two different wavelengths to quantify the level of proteolysis (Figure 3.14). At the rhodamine wavelength, the fluorescent spots corresponded to the non-cleaved peptides, where the fluorescein fluorescence was quenched by rhodamine, while at the fluorescein wavelength, the fluorescent spots corresponded to the cleaved peptides, where a disruption of the FRET system led to an increase in fluorescein intensity. The protease substrate specificity was accurately profiled by calculating the ratios of fluorescein/rhodamine for every feature, with the highest ratios of fluorescein/rhodamine corresponding to the preferred amino acids for a specific position within the tetrapeptide (Figure 3.14).

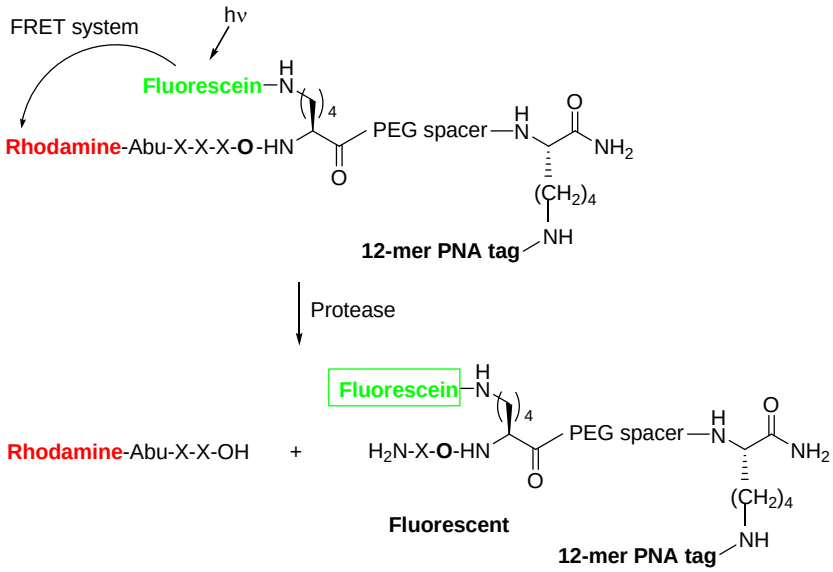


Figure 3.13: Example of a proteolytic cleavage of a PNA-encoded FRET-based peptide positional scanning library, where **O** = encoded amino acid and X = equimolar mixture of amino acids

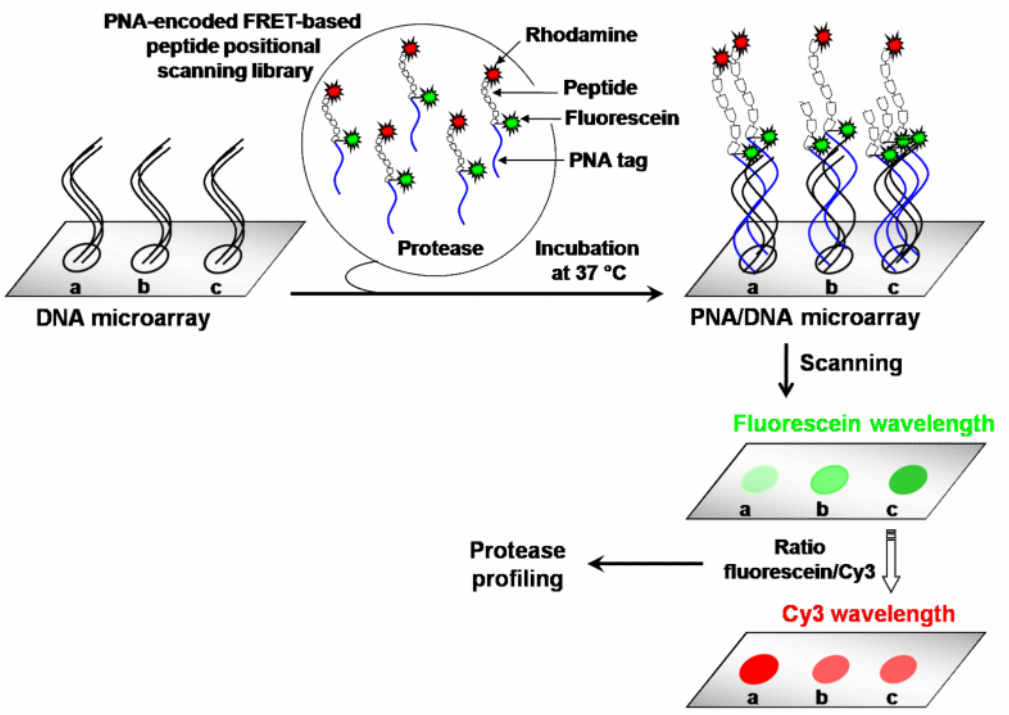


Figure 3.14: Protease profiling using a PNA-encoded FRET-based peptide positional scanning library

3.8.2 Experiments using the 625-member PNA-encoded positional scanning library (first sub-library)

3.8.2.1 Proof of solution-phase proteolytic cleavage

Before carrying out microarray-based proteolytic profiling, it was essential to verify that peptides from the 625-member PNA-encoded positional scanning library (first sub-library) could act as protease substrates. To do so, the first sub-library (100 µM final concentration) was assayed in solution with the serine protease subtilisin A (10 µM final concentration), and the fluorescein emission was monitored by spectrofluorometry, at an excitation wavelength of 480 nm and an emission wavelength of 520 nm. After 30 min incubation, a significant increase in fluorescein fluorescence was observed, proving that proteolytic cleavage occurred (Figure 3.15). Two peaks should have been observed at 520 nm (fluorescein emission) and at 585 nm (rhodamine emission). Although there was no peak at 585 nm, the fluorescent images displayed in Figure 3.8 confirmed the presence of rhodamine, and the “red” colour of all resin beads indicated that the beads were properly labelled with rhodamine. Moreover, in future experiments reported in section 3.8.2.3, the images

of the microarray obtained after hybridisation with the first sub-library (see Figure 3.18) showed fluorescent spots when using a rhodamine filter.

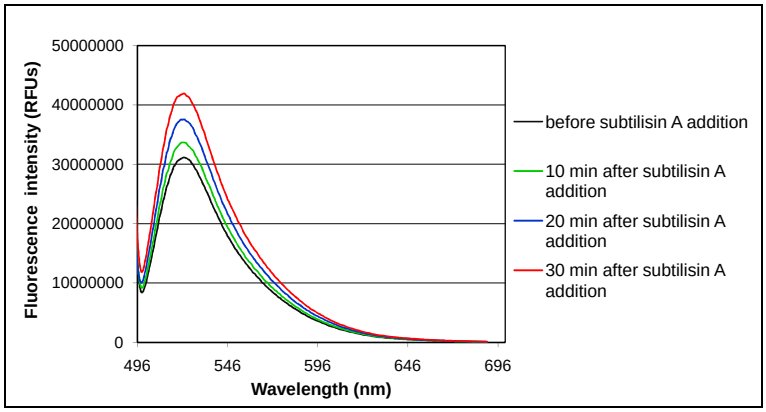


Figure 3.15: Fluorescence spectra obtained with the 625-member PNA-encoded positional scanning library (first sub-library) before and after incubation with subtilisin A

3.8.2.2 Preparation of DNA microarrays

Five different 3'-amino modified 12-mer DNAs (Table 3.3), each complementary to one of the five PNA tags (Table 3.1), with an additional spacer composed of six thymines (to separate the DNA/PNA from the microarray surface) were printed onto commercially available aldehyde-derivatised glass slides (see section 1.5.2) using a robotic microarrayer. To verify successful DNA printing on the surface, a 3'-amino-5'-fluorescein modified 12-mer DNA, whose sequence was not complementary to any of the five PNA tags, was also printed as a control. Each DNA oligomer was printed as five replicates (1 x 5 sub-arrays), and this pattern was duplicated across the microarray (Figure 3.16). In the following sections, all the positions of the printed DNAs were named according to the encoded amino acid (*e.g.* the names Ala, Ser, Lys, Glu and Pro corresponded to the DNAs complementary to the PNAs that encoded the amino acids Ala, Ser, Lys, Glu and Pro in the first position of the tetrapeptide respectively).

Table 3.3: Sequences of the modified 12-mer DNAs

Encoded amino acid	Modified 12-mer DNA (from 5' → 3')	
	3'-amino	3'-amino-5'-fluorescein
Ala	GATGATCAAGTC	
Ser	GATGATCAACCG	
Lys	GATGATCGAGTT	
Glu	GATGATCGATGA	
Pro	GATGATCGACAG	
Control		TTCCGTACAATG

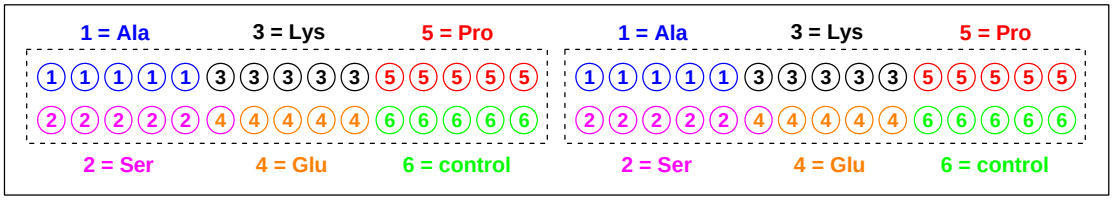


Figure 3.16: Schematic representation of the pattern used for DNA printing on a microarray

After DNA printing, the slides were washed and scanned using a fluorescein filter. Fluorescent spots where the fluorescein-labelled DNA (control) was printed were observed, confirming that the printing was successful.

3.8.2.3 Hybridisation onto DNA microarrays

To determine the best library concentration for hybridisation, the first sub-library was hybridised onto a DNA microarray at four different concentrations (5 µM, 500 nM, 50 nM and 5 nM). Following washing and scanning of each microarray using a rhodamine and a fluorescein filter, the intensities of the spots were analysed using the FIPS software (LaVision BioTech). Median values and standard errors were calculated for the five different features containing the same encoded amino acid, which was followed by the calculation of the average median values and standard errors of each encoded amino acid between the duplicated patterns (Figure 3.17). The greater the PNA hybridisation, the greater the rhodamine fluorescence was. A 5 µM concentration was found to be optimal, producing the highest (but not saturated) fluorescent signals. The choice of a lower concentration would have generated fluorescent signals too low such that it would have been more difficult to interpret the data.

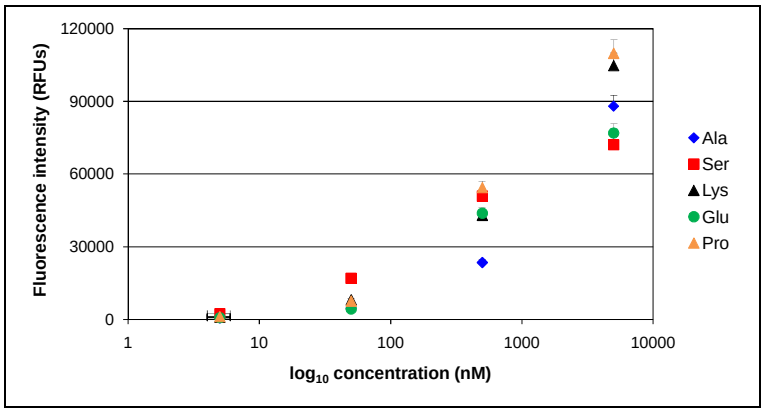


Figure 3.17: Correlation between rhodamine intensity and the first sub-library concentration (values correspond to the average of ten replicates)

Images of the microarray obtained after hybridisation with the first sub-library at a 5 µM concentration are displayed in Figure 3.18. At the rhodamine wavelength, it was confirmed that hybridisation was selective since no fluorescent signal was observed where the non-complementary fluorescein-labelled DNA (control) was printed and each PNA hybridised with its complementary DNA. All the replicates gave very similar fluorescent spots. Small differences in the rhodamine spot intensity between different encoded amino acids were observed (Figures 3.17 and 3.18), confirming that all PNAs did not hybridise with the same efficiency and that internal controls were necessary. At the fluorescein wavelength, the control fluorescence intensity was very high, and very weak fluorescent spots (close to the background intensity) were observed for the encoded amino acids since fluorescein was quenched by rhodamine (Figure 3.18).

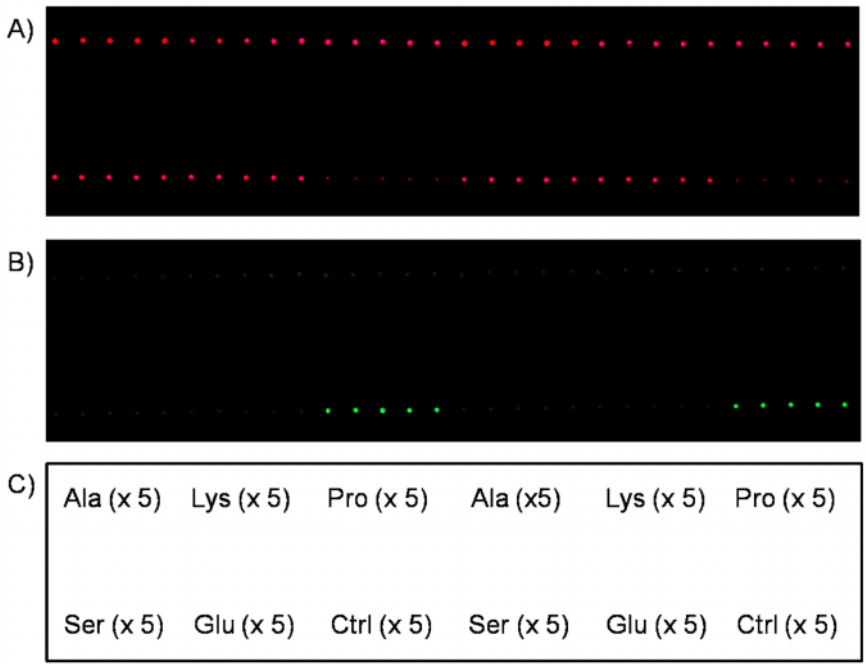


Figure 3.18: Images of the microarray hybridised with the first sub-library using A) a rhodamine filter and B) a fluorescein filter; C) schematic representation of the microarray pattern, where Ctrl is the fluorescein-labelled control

3.8.2.4 Enzymatic assays

All the enzymatic assays were performed following the procedure described in section 3.8.1. The data were reprocessed by calculating the ratio of fluorescein/rhodamine for each feature, and determining the median values and standard errors of the ratio of fluorescein/rhodamine for each encoded amino acid, which was followed by the calculation of the average median values and standard errors of each encoded amino acid between the duplicated patterns. The 625-member PNA-encoded positional scanning library (first sub-library) was used to profile the substrate specificities of three different proteases.

The first enzymatic assay was carried out using the aspartic endopeptidase pepsin. Pepsin, whose name was given by Schwann in 1836, was the first animal enzyme to be identified and is one of the main digestive proteases of the stomach (along with chymotrypsin and trypsin).[193, 225] Pepsin cleaves peptide bonds between large hydrophobic/aromatic amino acids, with the preferential cleavage sites being the P_1 and P_1' positions, where the large hydrophobic residues Phe and Leu are preferred. In general, pepsin has a very broad specificity (this is appropriate due to its role in digestion) and accepts a large variety of amino acids.[193] The images of the

microarray hybridised with the first sub-library following pepsin incubation in solution are shown in Figure 3.19. The use of the rhodamine wavelength confirmed that the hybridisation was selective, while new fluorescent spots corresponding to peptides cleaved by pepsin could be observed at the fluorescein wavelength. Since some fluorescent spots were close to the background intensity at the fluorescein wavelength (*e.g.* at the positions where the DNA that encoded the amino acid Ala was printed), the fluorescein-labelled control fluorescence intensity was increased up to saturation so that it was easier to observe the other fluorescent spots. If the same approach would have been applied to Figure 3.18, only small differences in the fluorescein spot intensity would have been observed between the different encoded amino acids, and the replicate intensity would have been very similar. In Figure 3.19, the fact that a proteolytic cleavage occurred was confirmed by the presence of important differences between the different encoded amino acid fluorescent signals at the fluorescein wavelength. If no peptides were recognised by pepsin, similar fluorescein spot intensities between the different encoded amino acids would have been observed. Following data reprocessing (see Appendix IV), a first profile was obtained, confirming the broad specificity of pepsin since all five residues were accepted and no particular amino acid was strongly preferred (Figure 3.20). It should be pointed out that the approach described here detected cleavage anywhere in the peptide, and the preferred amino acids (but not the consensus sequences) were identified at each position of the peptide.

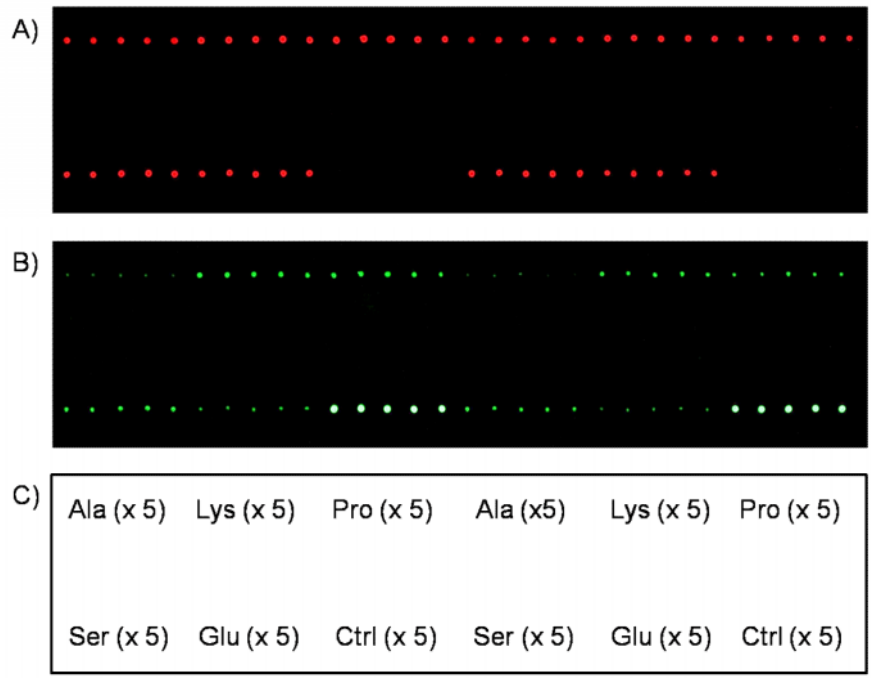


Figure 3.19: Images of the microarray hybridised with the first sub-library following pepsin incubation in solution using A) a rhodamine filter and B) a fluorescein filter; C) schematic representation of the microarray pattern, where Ctrl is the fluorescein-labelled control

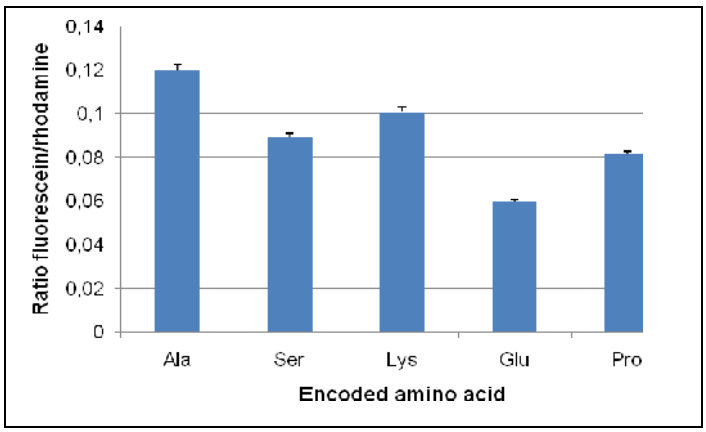


Figure 3.20: Profile of pepsin obtained using the 625-member PNA-encoded positional scanning library (first sub-library) (values correspond to the average of ten replicates)

A second enzymatic assay was carried out using the serine endopeptidase subtilisin A that was discovered in 1947 by Linderström-Lang and Ottensen.[226] Subtilisins are also non-specific proteases, P_1 and P_4 being the most important positions. In the P_1 position, large and non-β-branched hydrophobic amino acids are preferred but basic amino acids can be tolerated. Hydrophobic amino acids are strongly preferred in the P_4 position, the P_3 position accepts any amino acid and the P_2 position prefers Ala.[193]

Following data reprocessing (see Appendix V), the results confirmed that subtilisin A was not a highly specific protease (Figure 3.21) since polar, basic, acidic and hydrophobic residues were all tolerated (similar ratios for Ser, Lys, Glu and Pro). Interestingly the small hydrophobic residue Ala was clearly the most preferred amino acid, which agreed well with the known specificity of the P_2 position.[193]

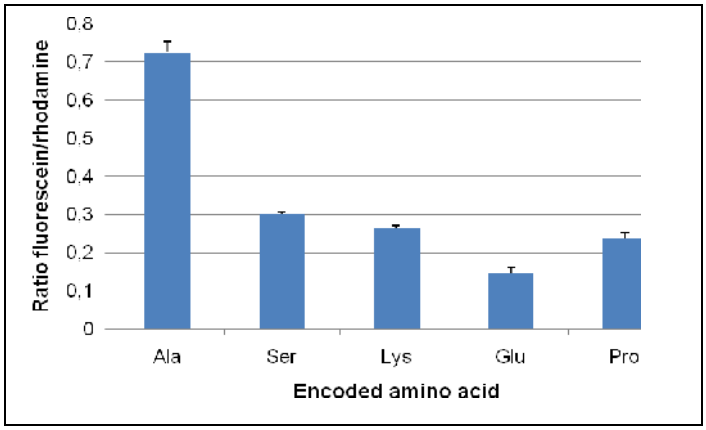


Figure 3.21: Profile of subtilisin A obtained using the 625-member PNA-encoded positional scanning library (first sub-library) (values correspond to the average of ten replicates)

The last enzymatic assay was performed using the serine endopeptidase trypsin that was discovered in 1876 by Kühne.[193] Trypsin is a digestive protease that strongly prefers to cleave peptide bonds on the carboxyl side of Lys and Arg. Large hydrophobic and positively charged amino acids are preferred in the P_1' and P_2' positions respectively.[193] Following data reprocessing (see Appendix VI), trypsin showed the expected preference for the basic amino acid Lys[193, 201] (Arg was not present in the library) but also a marked preference for Ala (Figure 3.22), which agreed particularly well with the results reported by Jakubke, who described the peptide sequence Gly-Pro-Ala-Lys-Leu-Ala-Ile-Gly as a good substrate for trypsin (cleavage after the residue Lys).[227] By looking at proteolytic cleavage anywhere in the peptide, the residue Lys that was known to play a crucial role in trypsin activity and the cooperative role of Ala were both identified in a single experiment. This strategy would represent a clear advantage when studying a protease whose substrate specificity is unknown.

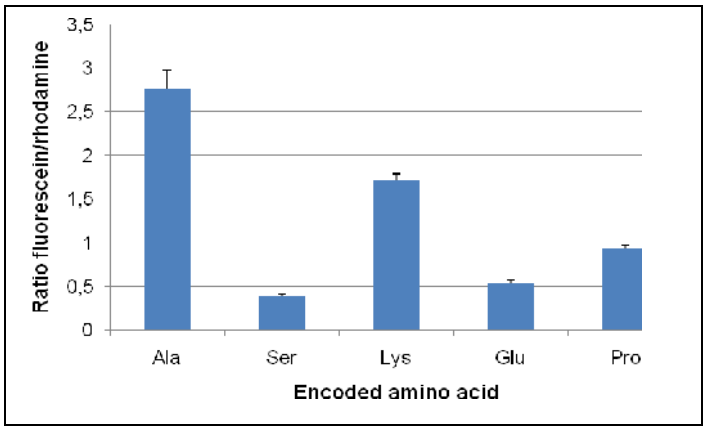


Figure 3.22: Profile of trypsin obtained using the 625-member PNA-encoded positional scanning library (first sub-library) (values correspond to the average of ten replicates)

3.8.3 Experiments using the 50,625-member PNA-encoded positional scanning library

3.8.3.1 Proof of solution-phase proteolytic cleavage

Each sub-library (50 µM final concentration) was separately assayed in solution with trypsin (20 nM final concentration), and the fluorescein emission was monitored by spectrofluorometry, at an excitation wavelength of 480 nm. Proteolytic cleavage for each sub-library was confirmed by the significant increase of the fluorescein intensity (Figure 3.23). Although the four sub-libraries had the same concentration (based on UV absorbance measurements), different profiles for the emission spectra were obtained. When the first, second and fourth sub-libraries were excited at 480 nm, two peaks were observed at 520 nm (fluorescein emission) and at 585 nm (rhodamine emission), with the peak at 520 nm being the most intense; for the third sub-library, both peaks had similar intensities. Since proteolytic cleavage was clearly observed in solution for the four sub-libraries, experiments were carried out using DNA microarrays.

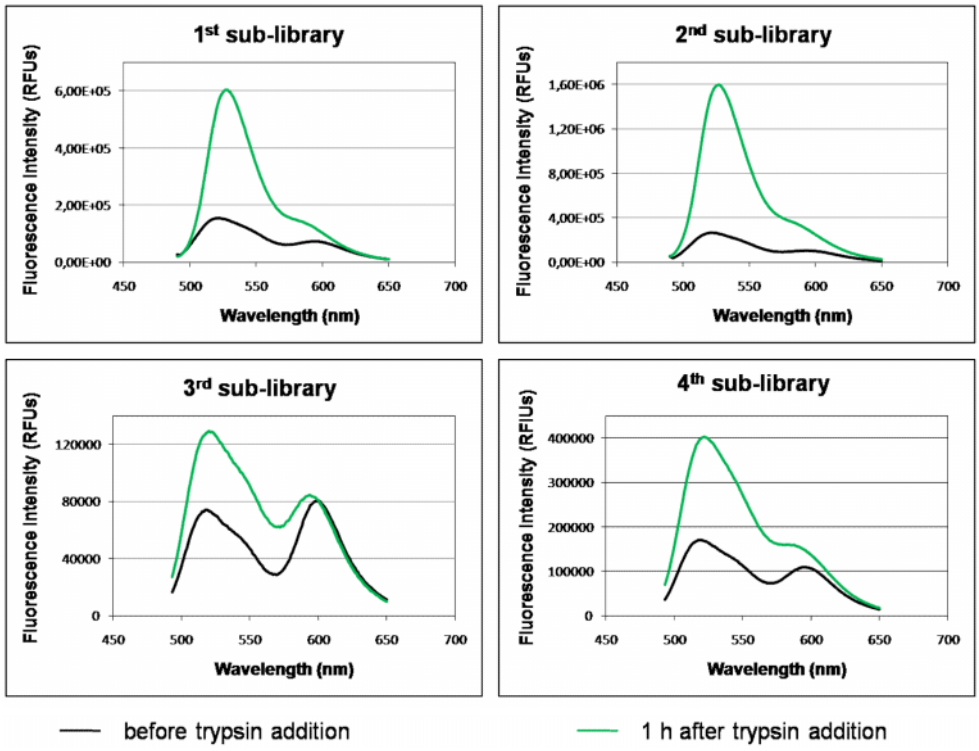


Figure 3.23: Fluorescence spectra obtained with the 50,625-member PNA-encoded positional scanning library before and after incubation with trypsin

3.8.3.2 Preparation of DNA microarrays

DNA microarrays were prepared following the procedure described in section 3.8.2.2, except that fifteen different 3'-amino modified 12-mer DNAs (Table 3.4), each complementary to one of the fifteen PNA tags (Table 3.2), were printed onto aldehyde-derivatised glass slides as five replicates (1 x 5 sub-arrays) following the pattern shown in Figure 3.24C.

Table 3.4: Sequences of the modified 12-mer DNAs (from 5' → 3')

	1st sub-library	2nd sub-library	3rd sub-library	4th sub-library
Ala	TAGA TAGA TAGA	ATGT TAGA TAGA	TGAT TAGA TAGA	AGTA TAGA TAGA
Arg	TAGA ATGT TAGA	ATGT ATGT TAGA	TGAT ATGT TAGA	AGTA ATGT TAGA
Gln	TAGA TGAT TAGA	ATGT TGAT TAGA	TGAT TGAT TAGA	AGTA TGAT TAGA
Glu	TAGA AGTA TAGA	ATGT AGTA TAGA	TGAT AGTA TAGA	AGTA TAGA ATGT
Gly	TAGA TAGA ATGT	ATGT TAGA ATGT	TGAT TAGA ATGT	AGTA ATGT ATGT
His	TAGA ATGT ATGT	ATGT ATGT ATGT	TGAT ATGT ATGT	AGTA TGAT ATGT
Ile	TAGA TGAT ATGT	ATGT TGAT ATGT	TGAT AGTA ATGT	AGTA AGTA ATGT
Lys	TAGA AGTA ATGT	ATGT AGTA ATGT	TGAT TAGA TGAT	AGTA TAGA TGAT
Phe	TAGA TAGA TGAT	ATGT TAGA TGAT	TGAT ATGT TGAT	AGTA ATGT TGAT
Pro	TAGA ATGT TGAT	ATGT TGAT TGAT	TGAT TGAT TGAT	AGTA TGAT TGAT
Ser	TAGA TGAT TGAT	ATGT AGTA TGAT	TGAT AGTA TGAT	AGTA AGTA TGAT
Thr	TAGA AGTA TGAT	ATGT TAGA AGTA	TGAT TAGA AGTA	AGTA TAGA AGTA
Trp	TAGA TAGA AGTA	ATGT ATGT AGTA	TGAT ATGT AGTA	AGTA ATGT AGTA
Tyr	TAGA ATGT AGTA	ATGT TGAT AGTA	TGAT TGAT AGTA	AGTA TGAT AGTA
Val	TAGA TGAT AGTA	ATGT AGTA AGTA	TGAT AGTA AGTA	AGTA AGTA AGTA

3.8.3.3 Enzymatic assays

Compared to the 625-member PNA-encoded library, a higher library concentration was required to hybridise the 50,625-member PNA-encoded positional scanning library (all four sub-libraries) onto the DNA microarray. Since a 25 µM final library concentration produced low fluorescent signals, a 50 µM final library concentration was used. Images of the microarray obtained after hybridisation are displayed in Figure 3.24. At the rhodamine wavelength, the selectivity of the hybridisation was confirmed, while at the fluorescein wavelength, low fluorescent spots were observed due to the quenching of fluorescein by rhodamine.

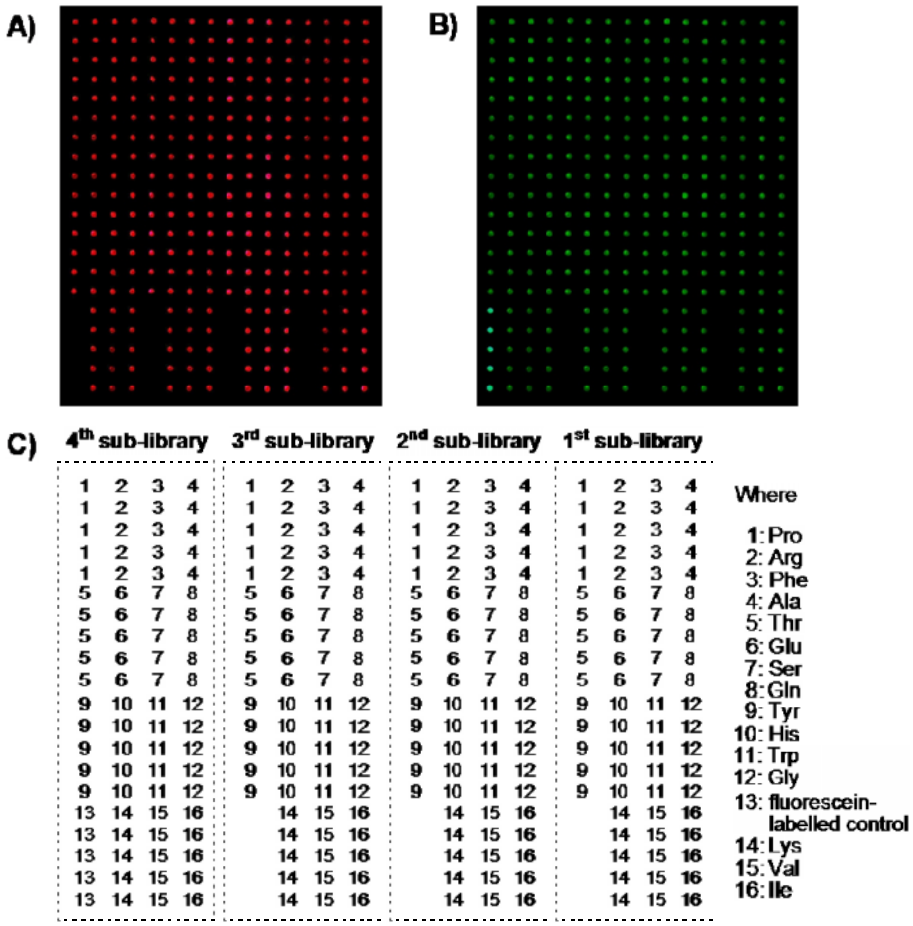


Figure 3.24: Images of the microarray hybridised with the 50,625-member PNA-encoded positional scanning library using A) a rhodamine filter and B) a fluorescein filter; C) schematic representation of the microarray pattern

The first enzymatic assays were carried out by incubating the 50,625-member PNA-encoded positional scanning library (all four sub-libraries) with trypsin for 2 h at 37

°C. Images of the microarray hybridised with the library following trypsin incubation are shown in Figure 3.25. At the rhodamine wavelength, each PNA tag of the four sub-libraries hybridised with its complementary DNA, generating fluorescent spots with similar intensities between the replicates but also between the different encoded amino acids since the PNA tags were designed to have similar T_m's. However, at the fluorescein wavelength, instead of having spots with different intensities according to the peptide cleavage rate, all fluorescent spots had virtually identical intensities.

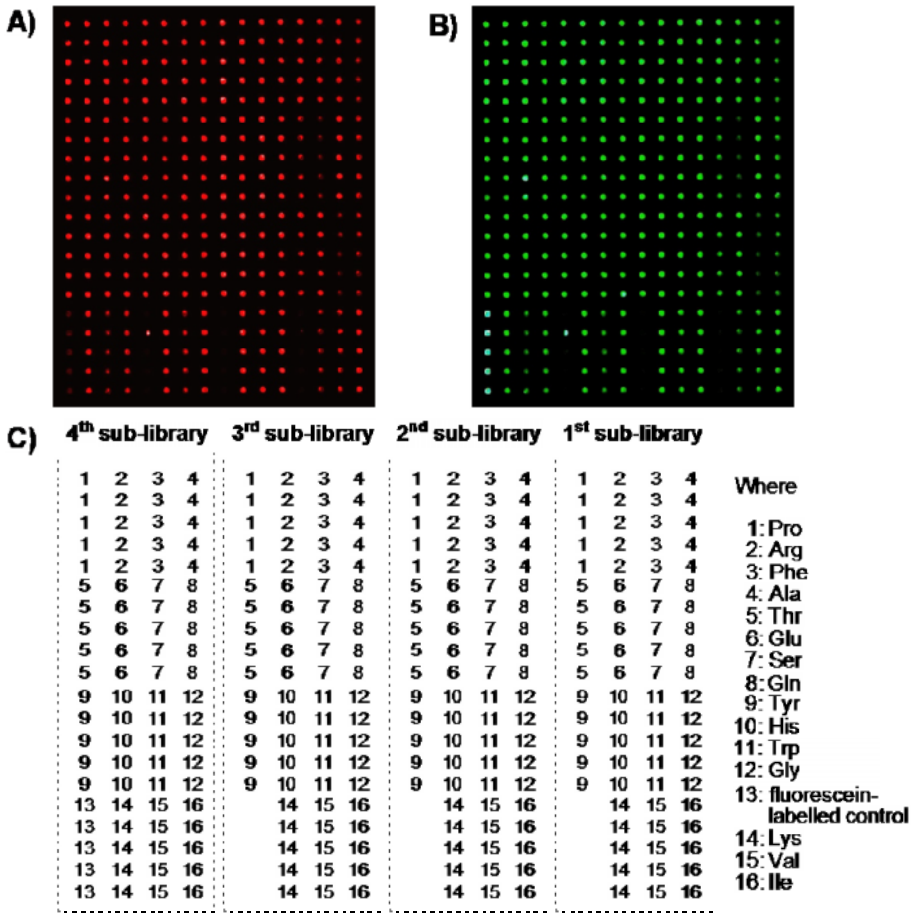


Figure 3.25: Images of the microarray hybridised with the 50,625-member PNA-encoded positional scanning library following trypsin addition using A) a rhodamine filter and B) a fluorescein filter; C) schematic representation of the microarray pattern

The data were reprocessed by calculating the ratio of fluorescein/rhodamine for each feature, and determining the median values and standard errors of the ratio of fluorescein/rhodamine for each encoded amino (see Appendix VII). The profile obtained showed no detectable variation in substrate specificity (Figure 3.26). A

clear preference for positively charged (Lys, Arg) and hydrophobic residues (Ala) should have been observed (see section 3.8.2.4). Different assay conditions were tested, varying the concentration of trypsin (from 1 nM to 2 µM) and the time of incubation (from 30 min to 16 h), or by carrying out on-chip assays (a fresh solution of trypsin (450 nM) was added every hour onto the PNA/DNA microarray over a period of 6 h) but no difference was observed.

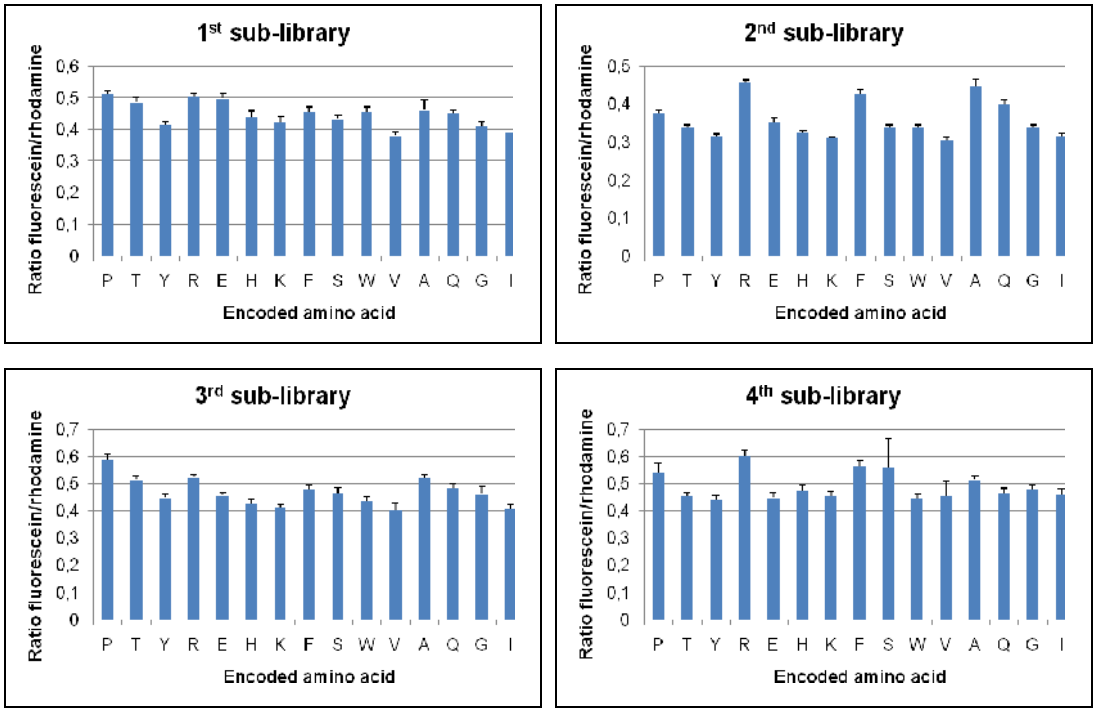


Figure 3.26: Profiles of trypsin obtained using the 50,625-member PNA-encoded positional scanning library (values correspond to the average of five replicates)

Enzymatic assays were carried out using two other enzymes, the cysteine protease chymopapain (Figure 3.27, see Appendix VIII for data reprocessing) and the aspartic protease pepsin (Figure 3.28, see Appendix IX for data reprocessing) but once again, very similar ratios of fluorescein/rhodamine for all the different encoded amino acids were observed. A broad specificity was expected for pepsin but chymopapain prefers the amino acids Ala, Pro, Val and Lys in positions P_4 to P_1.[228] The fluorescent spots observed at both rhodamine and fluorescein wavelengths as well as the results obtained by spectrofluorometry (see section 3.8.3.1) proved that a FRET system was present in the library and that proteolytic cleavage occurred in solution. However, once the peptides were immobilised onto the microarray, it was not possible to detect

any cleavage. The only differences between the 625- and the 50,625-member PNA-encoded positional scanning libraries were the number of library members and the PNA tag sequences. Each microarray spot would contain 125 or 3,375 different peptides (625- or 50,625-member library respectively). One hypothesis to explain the results could be that if too many peptides were hybridised at the same location then there could be some interactions between the two dyes fluorescein (cleaved peptides) and rhodamine (non-cleaved peptides), thus affecting the fluorescence intensity.

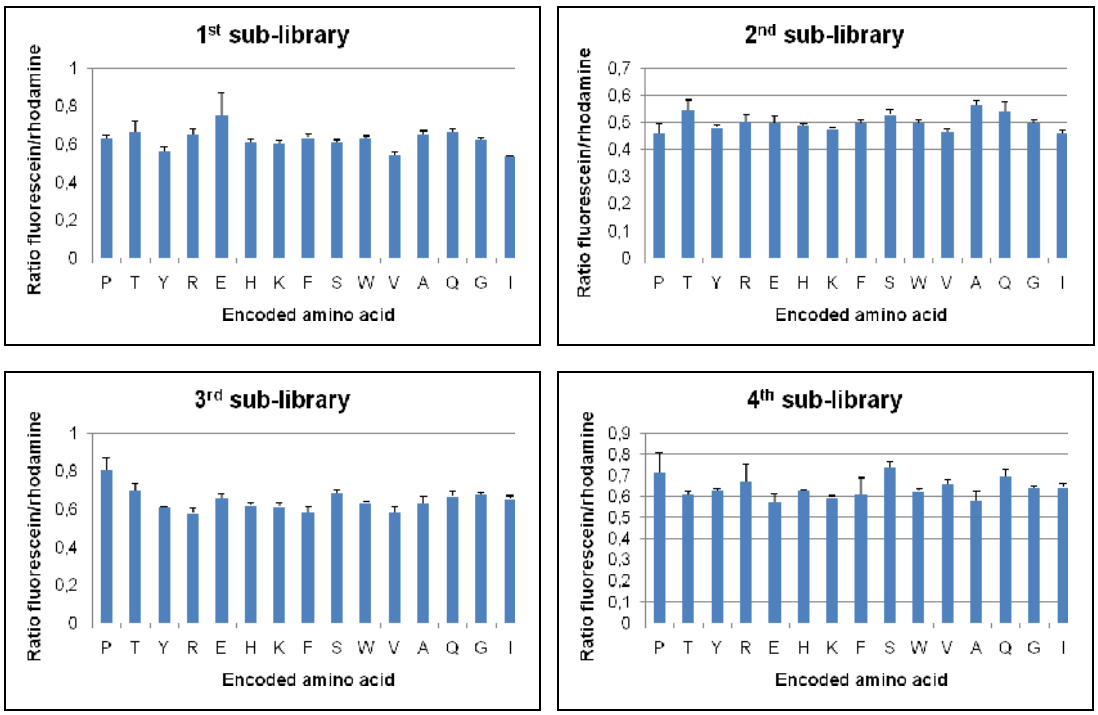


Figure 3.27: Profiles of chymopapain obtained using the 50,625-member PNA-encoded positional scanning library (values correspond to the average of five replicates)

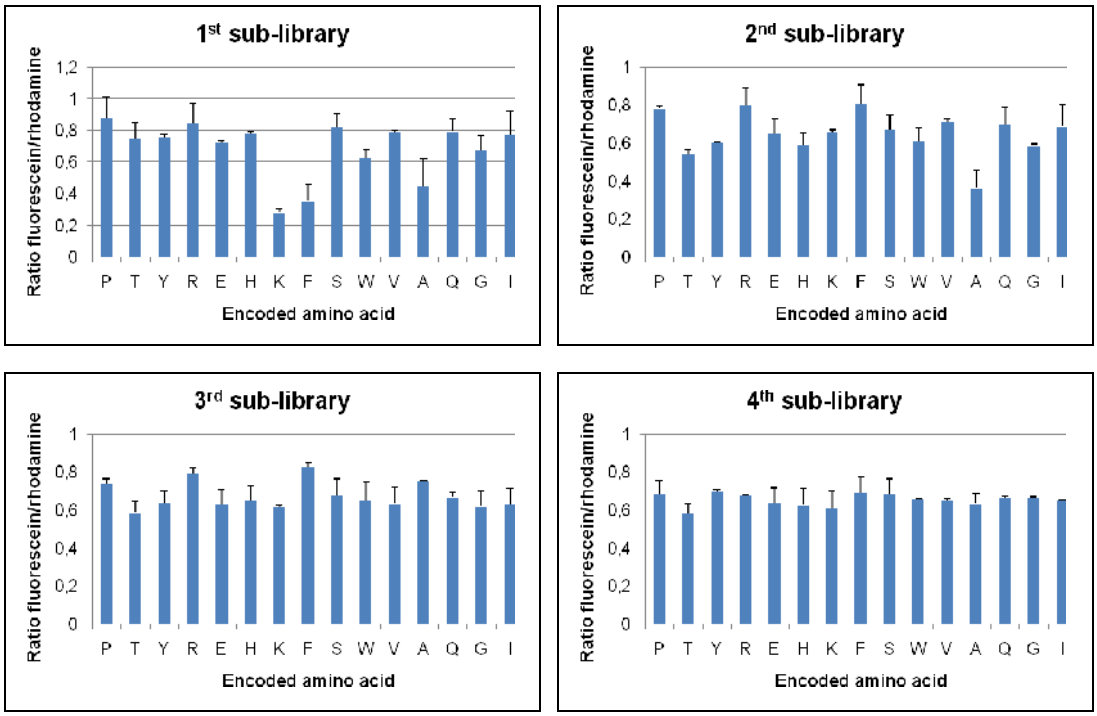


Figure 3.28: Profiles of pepsin obtained using the 50,625-member PNA-encoded positional scanning library (values correspond to the average of five replicates)

3.9 Conclusion

Two PNA-encoded positional scanning libraries of FRET-based tetrapeptides (625 or 50,625 members) were successfully synthesised by solid phase synthesis. The "prime" and "non-prime side" substrate specificities of three different proteases (pepsin, subtilisin A and trypsin) were easily and rapidly profiled using the 625-member positional scanning library (first sub-library), with assays performed in solution using minimal amounts of library and protease. Both cleaved and non-cleaved peptides were hybridised onto a DNA microarray, with just five PNA tags allowing the rapid deconvolution and identification of the preferred amino acids for the first position in the tetrapeptide. This strategy was then applied to a larger 50,625-member PNA-encoded positional scanning library. The first enzymatic assays were performed using trypsin, with cleavage demonstrated. However, the interpretation of the microarray data revealed a "non-specific" profile, with trypsin accepting all fifteen amino acids equally! This result disagreed with the known specificity of trypsin, and the same conclusions were observed using two other

proteases (chymopapain and pepsin). It was supposed that the immobilisation of a large number of library members onto a microarray within a few locations could be responsible for interactions between the two dyes fluorescein and rhodamine, affecting spot fluorescence and consequently the interpretation of the results.

Chapter 4: Experimental part

4.1 General section

4.1.1 General information

All solvents and reagents were obtained from commercial suppliers and used without purification, unless otherwise stated.

NMR spectra were recorded on Bruker ARX-250, DPX-360 and AVA-800 spectrometers operating at 250 MHz, 360 MHz and 800 MHz for ^{1}H and 62.5 MHz, 90 MHz and 200 MHz for ^{13}C spectra respectively, in the solvents indicated at 298 K. ^{31}P NMR spectra were recorded on a Bruker ARX-250 at 101 MHz. All chemical shifts (δ) are quoted in ppm using the residual hydro solvent as the internal standard or the ^{13}C natural abundance of the deuterated solvent. Coupling constants (*J*) are reported in Hz.

Infrared (IR) spectra were obtained on a Fourier transform IR Bruker Tensor 27 Spectrometer (FTS) fitted with a Specac single reflection diamond attenuated total reflection (ATR) Golden Gate. All samples were run neat, frequencies are reported in cm^{-1} and only frequencies corresponding to significant functional groups are reported.

Aluminium backed silica plates (Merck silica gel 60 F_{254}) were used for thin layer chromatography (TLC) to monitor solution-phase reactions. TLC visualisation was carried out using short wavelength ultraviolet (UV) light at 254 nm. Flash column chromatography was performed on silica gel using Keiselgel 60, 230-400 mesh, as supplied by Merck.

Microwave assisted peptide syntheses were performed using a CEM Liberty™ microwave peptide synthesiser. All amino acids used were of L configuration, unless specified. MALDI-TOF MS analyses were performed on an Applied Biosystems Voyager DESTR MALDI-TOF mass spectrometer. Sinapic acid was used as a matrix and positive ion mass spectra are reported.

HPLC spectra were obtained using Agilent Technologies HP1100 Chemstations. Solvents used were (A) H_2O/0.1% TFA and (B) CH_3CN/0.04% TFA, (C) H_2O/0.1%

formic acid and (D) MeOH/0.1% formic acid, at a flow rate of 1 mL/min. The columns used were a Discovery C18 (column 1) from Supelco (50 mm x 4.6 mm, 5 µm), a Luna C18 (column 2) from Phenomenex (150 mm x 4.6 mm, 5 µm) and a Gemini C18, 110 Å (column 3) from Phenomenex (100 mm x 4.6 mm, 5 µm). The following methods were used:

- Method 1 [column 1, eluents (A) and (B)]: 95% to 5% (A) over 6 min, 5% (A) for 2 min
- Method 2 [column 1, eluents (A) and (B)]: 90% to 10% (A) over 3 min, 10% (A) for 1 min
- Method 3 [column 2, eluents (C) and (D)]: 100% (C) for 5 min, 0% to 60% (D) over 10 min, 100% (D) for 5 min
- Method 4 [column 3, eluents (C) and (D)]: 95% to 5% (C) over 10 min

Analytical HPLC/evaporative light scattering detector (ELSD) spectra were obtained with the same system (using the same eluents, methods and flow rates) coupled to a Polymer Lab 100 ES ELSD.

Electrospray ionisation (ESI)/MS analyses were performed on an Agilent Technologies LC/MSD Series 1100 quadrupole mass spectrometer (QMS), equipped with an electrospray ion source. Major peaks are reported with percentage intensities of the base peak.

Preparative reverse phase-HPLC purifications were performed on an Agilent Technologies HP1100 Chemstation eluting with (A) H_2O/0.1% TFA and (B) CH_3CN/0.1% TFA on a Waters X-Terra RP18 preparative column (150 mm x 19 mm, 5 µm), at a flow rate of 5 mL/min. The following gradient was used: 100% (A) to 40% (B) over 5 min, 40% to 70% (B) over 20 min, 70% to 100% (B) over 4 min, detection by UV at 214, 254 and 440 nm.

Absorption spectra were recorded on an 8453 Agilent UV-Visible spectrophotometer. Fluorescence measurements were performed on a Jobin Yvon Spex Fluoromax spectrofluorometer or a FS900 spectrofluorometer.

Microarray printing was performed using a Genetix QArrayMini (Genetix, UK) contact printer, and hybridisation was carried out using the GeneMachines Hyb4 automated station (Genomic Solutions, UK). Microarrays were scanned with a white

light scanner (BioAnalyser 4F, LaVision BioTech, Germany). Aldehyde-derivatised microscope slides and CodeLink™ activated glass slides were obtained from Genetix (UK) and Amersham Biosciences (UK) respectively.

4.1.2 General experimental procedures

Calculation of theoretical loading

Theoretical loading of a resin after a reaction was calculated using the equation:

New loading (mmol/g) = [old loading/(1 + (old loading x mass added x 10^{-3})]

Where: "old loading" is the loading of the starting resin (mmol/g) and "mass added" is the mass of compound added onto the resin

Quantitative Fmoc test[229]

To a known mass of resin (*e.g.* 5 mg), pre-swelled in DMF for 5 min, was added a solution of 20% piperidine in DMF (1 mL), and the mixture was shaken for 5 min. The solution cleavage was added to a 10 mL volumetric flask and the volume was made up to 10 mL with a solution of 20 % piperidine in DMF. The absorbance at 302 nm of the liberated dibenzofulvene-piperidine adduct was measured twice against a blank solution of 20 % piperidine in DMF and the average value was used. The loading was calculating using the equation:

Loading (mmol/g) = [(A_{302} x V)/(ε_{302} x m)] x 10^3

Where A_{302} is the recorded absorbance at 302 nm, V is the diluted volume (10 mL), ε_{302} is the extinction coefficient of the dibenzofulvene-piperidine adduct (7800 M^{-1} cm^{-1}) and m is the mass of resin (mg).

Qualitative nynhidrin test[220]

To a few resin beads placed in a small test tube were added 3 drops of reagent A and 1 drop of reagent B. The mixture was heated at 100 °C for 5 min. The presence of resin bound free amine was indicated by a blue colour.

Reagent A: Solution 1: Potassium cyanide (65 mg, 1 mmol) was dissolved in H_2O (100 mL). An aliquot of this solution (2 mL) was diluted with freshly distilled pyridine (98 mL). Solution 2: Phenol (40 g, 420 mmol) was dissolved in absolute EtOH (10 mL). Solutions 1 and 2 were mixed together to give reagent A.

Reagent B: Ninhydrin (2.5 g, 14 mmol) was dissolved in absolute EtOH (50 mL).

Quantitative ninhydrin test[230]

To a known mass of resin (*e.g.* 5 mg) placed in a small test tube was added 6 drops of reagent A (see above) and 2 drops of reagent B (see above). The mixture was heated at 100 °C for 5 min. The test tube was placed in a cold water bath, 60% aqueous ethanol (2 mL) was added and mixed thoroughly. The resin was removed by filtration through cotton wool and the deep blue filtrate collected in a 25 mL volumetric flask. The resin was washed with a solution of Et_4NCl (0.5 M in DCM, 2 x 0.5 mL) and the sample made up to 25 mL with 60% aqueous ethanol. The absorbance at 570 nm was measured twice against a blank solution and the average value was used. The level of amine present was calculated using the equation:

Amount of primary amine residues (mmol/g) = $[(A_{570} \times V)/(\varepsilon_{570} \times m)] \times 10^3$

Where A_{570} is the recorded absorbance at 570 nm, V is the final volume (mL), ε_{570} is an extinction coefficient suitable for most peptides (1.5×10^4 M^{-1} cm^{-1}) and m is the mass of resin (mg).

Qualitative chloranil test[231]

To a few resin beads placed in a small test tube was added one drop of 2% acetaldehyde in DMF followed by one drop of 2% *p*-chloranil in DMF. The resin was allowed to stand for 5 min. Dark blue stained beads indicated the presence of secondary amines.

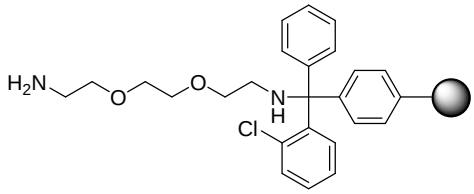
Fmoc deprotection

To a resin (*e.g.* 1 g), pre-swelled in DMF for 10 min, was added a solution of 20% piperidine in DMF (20 mL), and the mixture was shaken for 10 min. The resin was washed with DMF (3 x 25 mL) and the deprotection procedure was repeated. The resin was washed with DMF (3 x 25 mL) and DCM (3 x 25 mL).

4.2 Experimental for chapter 2

4.2.1 Fmoc solid-phase peptide synthesis

Preparation of amino PEG-trityl linker polystyrene resin (2.16)

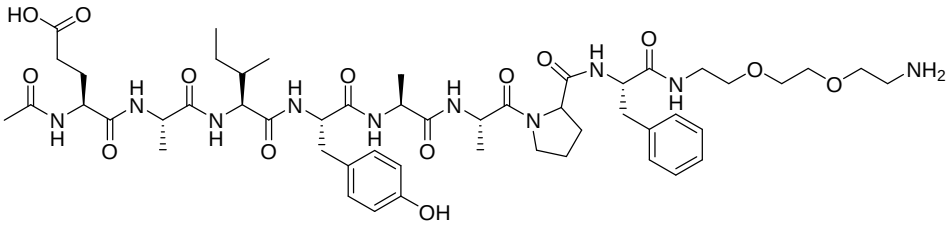


2-[2-(2-aminoethoxy)ethoxy]ethanamine (2.2 mL, 15 mmol, 10 eq) was dissolved in dry THF (10 mL) and added to 2-chlorotrityl chloride resin (1% divinylbenzene, loading 1.5 mmol/g, 1 g, 1.5 mmol, 1 eq) and the reaction was stirred for 16 h. The resin was washed with DCM/MeOH/DIPEA (17:2:1, 3 x 3 mL), DCM (3 x 3 mL), DMF (2 x 3 mL) and DCM (2 x 3 mL). The title yellow resin **2.16** was then dried *in vacuo* at 40 °C for 16 h.

Loading: determined by a quantitative ninhydrin test, 0.63 mmol/g

IR ν (cm^{-1}): 3055 (m, ν_{N-H}), 1667 (m, ν_{N-H}), 1105 (s, ν_{C-O})

Synthesis of Ac-Glu-Ala-Ile-Tyr-Ala-Ala-Pro-Phe-PEG-NH$_2$ (2.11)



Peptide **2.11** was synthesised using a CEM Liberty™ microwave peptide synthesiser. Fmoc deprotections were performed with a solution of 20% piperidine in DMF (2 cycles of 7 min at 25 °C) under N_2 bubbling. For every amino acid coupling, Fmoc-AA-OH (0.32 mmol, 5 eq) was dissolved in DMF (0.2 M) and added to the previously prepared resin **2.16** (loading 0.63 mmol/g, 100 mg, 0.063 mmol, 1 eq) followed by the addition of a solution of HBTU/HOBt (0.32 mmol, 5 eq) in DMF (0.25 M) and then a solution of 20 % DIPEA in *N*-methylpyrrolidinone (NMP) (0.63 mmol, 10 eq). A double coupling of 20 min, at 60 °C under microwave irradiation

and N_2 bubbling, was then carried out. Between amino acid couplings and Fmoc deprotections, the resin was automatically washed with DMF. At the end of the peptide synthesis, the resin was transferred to a solid phase extraction tube, washed with DMF (3 x 3 mL) and DCM (3 x 3 mL), capped with a solution of acetic anhydride and pyridine (50/50, 2 mL) for 1.5 h followed by washings with DMF (3 x 3 mL), DCM (3 x 3 mL), MeOH (2 x 3 mL) and DCM (2 x 3 mL). The completion of the capping was verified by the Kaiser test (see section 4.1.2). Finally, the peptide was cleaved from the resin using a TFA/DCM/TIS (50/45/5) solution (2 mL) for 2 h and precipitated with cold diethyl ether (10 mL). The residue was collected by filtration, dissolved in CH_3CN/H_2O, lyophilised and purified by reverse phase preparative HPLC to afford **2.11** as a white solid (52 mg, 78%).

HPLC (method 1): t_R = 4.9 min (100% purity, ELSD)

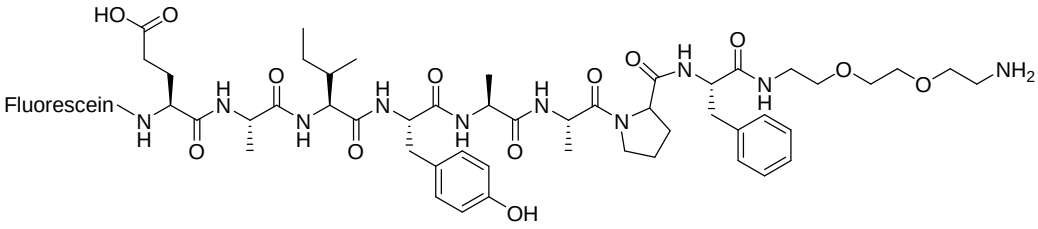
MALDI-TOF$^+$/MS: m/z (%) 1052.7 $[M+H]^+$ (100)

^{1}H NMR (800 MHz, D_2O): δ_H 7.34 (2H, t, *J* = 7.5 Hz, ArH-Phe), 7.32-7.18 (3H, m, ArH-Phe), 7.10-7.03 (2H, m, ArH-Tyr), 6.79-6.74 (2H, m, ArH-Tyr), 4.58-4.51 (1H, m, H_α-Tyr), 4.49-4.42 (2H, m, H_α-Phe and H_α-Ala), 4.35 (1H, dd, *J* = 8.5, 5.6 Hz, H_α-Pro), 4.31-4.18 (3H, m, H_α-Glu and 2 x H_α-Ala), 4.15-3.96 (1H, m, H_α-Ile), 3.75 (2H, td, *J* = 10.0, 7.0 Hz, H_δ-Pro), 3.72-3.68 (2H, m, CH_2O), 3.65-3.52 (4H, m, CH_2O), 3.48-3.36 (2H, m, CH_2O), 3.36-3.18 (2H, m, C*H*$_2$NH), 3.18-3.13 (2H, m, C*H*$_2$NH$_2$), 3.11-2.97 (4H, m, H_β-Phe and H_β-Tyr), 2.92-2.86 and 1.75-1.58 (1H, m, H_β-Ile), 2.50-2.40 (2H, m, H_γ-Glu), 2.24-2.15 and 1.84-1.75 (2H, m, H_β-Pro), 2.10-1.87 (4H, m, H_β-Glu and H_γ-Pro), 2.00 (3H, s, CH_3-Ac), 1.38-1.00 (2H, m, H_γ-Ile), 1.31 (3H, d, *J* = 7.1 Hz, CH_3-Ala), 1.27 (6H, t, *J* = 7.7 Hz, CH_3-Ala), 0.92-0.72 (6H, m, CH_3-Ile)

^{13}C NMR (200 MHz, D_2O): δ_C 179.6 (COO-Glu), 177.2 (CO-Ala), 177.0 (CO-Ac), 176.4 (CO-Pro), 176.2 (CO-Ala), 176.0 (CO-Glu), 175.8 (CO-Ala), 175.7 (CO-Ile), 175.5 (CO-Phe), 174.9 (CO-Tyr), 157.1 (COH-Tyr), 138.8 (C-Phe), 133.2 (2 x ArH-Tyr), 131.9 (2 x ArH-Phe), 131.4 (2 x ArH-Phe), 130.6 (C-Tyr), 129.9 (ArH-Phe), 118.1 (2 x ArH-Tyr), 72.2 (CH_2O), 72.1 (CH_2O), 71.3 (CH_2O), 69.1 (CH_2O), 63.1 (C_α-Pro), 61.0 (C_α-Ile), 57.9 (C_α-Phe), 57.5 (C_α-Tyr), 55.6 (C_α-Glu), 52.3 (C_α-Ala), 51.6 (C_α-Ala), 50.4 (C_α-Ala), 50.4 (C_δ-Pro), 41.8 (CH_2NH_2), 41.4 (CH_2NH), 39.7 (C_β-Phe), 38.9 (C_β-Tyr), 38.8 (C_β-Ile), 32.6 (C_γ-Glu), 31.9 (C_β-Pro), 29.0 (C_β-Glu),

27.3 (C_γ-Pro), 27.1 (C_γ-Ile), 24.3 (CH_3-Ac), 19.5 (CH_3-Ala), 18.9 (CH_3-Ala), 18.0 (CH_3-Ala), 17.3 (CH_3-Ile), 12.8 (CH_3-Ile)

Synthesis of Fluorescein-Glu-Ala-Ile-Tyr-Ala-Ala-Pro-Phe-PEG-NH_2 (2.13)[151]

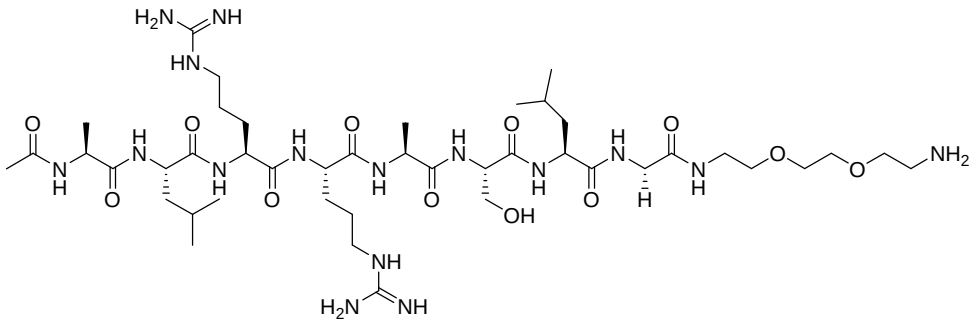


Peptide **2.13** was prepared as for peptide **2.11** but capped with 5(6)-carboxyfluorescein (fluorescein) (120 mg, 0.32 mmol, 5 eq) for 3 h (3 couplings) followed by washing with 20% piperidine in DMF (2 cycles of 10 min, 2 x 3 mL), DMF (2 x 3 mL) and DCM (2 x 3 mL). After purification by reverse phase preparative HPLC, the labelled peptide **2.13** was obtained as a yellow solid (52 mg, 60%).

HPLC (method 1): t_R = 5.5 min (100% purity, 440 nm)

MALDI-TOF$^+$/MS: m/z (%) 1369.0 $[M+H]^+$ (100)

Synthesis of Ac-Ala-Leu-Arg-Arg-Ala-Ser-Leu-Gly-PEG-NH_2 (2.9)



Peptide **2.9** was synthesised manually at 25 °C. The appropriate Fmoc amino acid (0.35 mmol, 5 eq), HBTU (133 mg, 0.35 mmol, 5 eq) and HOBt (47 mg, 0.35 mmol, 5 eq) were dissolved in DMF (1.4 mL) followed by addition of DIPEA (122 µL, 0.7 mmol, 10 eq). The resulting solution was mixed for 10 s before being added to the resin **2.16** (loading 0.63 mmol/g, 106 mg, 0.07 mmol, 1 eq) pre-swollen in DMF. The

mixture was shaken for 1 h. The completion of each coupling was verified by the Kaiser test (see section 4.1.2). If needed, an extended cycle was repeated until completion of the reaction. Fmoc deprotection was performed as described in section 4.1.2. After the last amino acid coupling, the resin was capped using acetic anhydride and pyridine (50/50, 2 mL) for 1.5 h. Finally, the peptide was cleaved from the resin using Reagent K (TFA/H_2O/Phenol/Thioanisole/EDT: 82.5/5/5/5/2.5, 2 mL) for 6 h and precipitated with cold diethyl ether (10 mL). The residue was collected by filtration, dissolved in CH_3CN/H_2O, lyophilised and purified by reverse phase preparative HPLC to afford **2.9** as an off-white solid (16 mg, 23%).

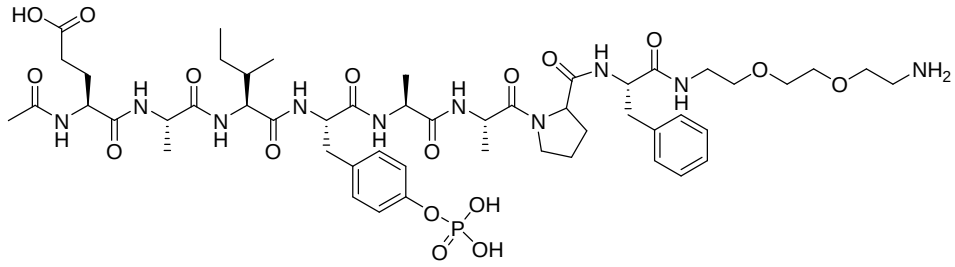
HPLC (method 1): t_R = 3.0 min (100% purity, ELSD)

MALDI-TOF$^+$/MS: m/z (%) 1015.0 $[M+H]^+$ (100)

^{1}H NMR (800 MHz, D_2O): δ_H 4.44 (1H, t, *J* = 5.7 Hz, H_α-Ser), 4.38 (1H, dd, *J* = 10.1, 4.7 Hz, H_α-Leu), 4.34-4.29 (4H, m, 2 x H_α-Arg and H_α-Ala and H_α-Leu), 4.22 (1H, q, *J* = 7.2 Hz, H_α-Ala), 3.90 (2H, d, *J* = 1.8 Hz, H_α-Gly), 3.89-3.84 (2H, m, H_β-Ser), 3.76 (2H, t, *J* = 5.1 Hz, CH_2O), 3.71 (4H, s, CH_2O), 3.64 (2H, t, *J* = 5.5 Hz, CH_2O), 3.45-3.39 (2H, m, C*H*$_2$NH), 3.23-3.17 (6H, m, 2 x H_δ-Arg + C*H*$_2$$NH_2$), 2.01 (3H, s, CH_3-Ac), 1.89-1.81 (2H, m, H_β-Arg), 1.80-1.73 (2H, m, H_β-Arg), 1.72-1.53 (10H, m, 2 x H_γ-Arg + 2 x H_β-Leu + 2 x H_γ-Leu), 1.41 (3H, d, *J* = 7.2 Hz, CH_3-Ala), 1.35 (3H, d, *J* = 7.2 Hz, CH_3-Ala), 0.94 (6H, t, *J* = 5.9 Hz, CH_3-Leu), 0.88 (6H, dd, *J* = 6.1, 4.5 Hz, CH_3-Leu)

^{13}C NMR (200 MHz, D_2O): δ_C 178.3 (CO-Ala), 177.8 (CO-Leu), 177.7 (CO-Leu), 177.4 (CO-Ala), 177.0 (CO-Ac), 176.2 (CO-Arg), 176.1 (CO-Arg), 174.6 (CO-Ser), 173.9 (CO-Gly), 159.4 (2 x C=NH-Arg), 72.3 (CH_2O), 72.2 (CH_2O), 71.5 (CH_2O), 69.1 (CH_2O), 63.6 (C_β-Ser), 58.1 (C_α-Ser), 56.1 (C_α-Arg), 56.0 (C_α-Arg), 55.3 (C_α-Leu), 55.1 (C_α-Leu), 52.7 (C_α-Ala), 52.6 (C_α-Ala), 45.2 (C_α-Gly), 43.2 (C_δ-Arg), 43.2 (C_δ-Arg), 42.3 (C_β-Leu), 42.2 (C_β-Leu), 41.8 (CH_2NH_2), 41.6 (CH_2NH), 30.8 (C_β-Arg), 30.5 (C_β-Arg), 27.1 (C_γ-Arg), 27.1 (C_γ-Arg), 27.0 (C_γ-Leu), 27.0 (C_γ-Leu), 24.9 (CH_3-Leu), 24.7 (CH_3-Leu), 24.2 (CH_3-Ac), 23.5 (CH_3-Leu), 23.3 (CH_3-Leu), 19.1 (CH_3-Ala), 19.0 (CH_3-Ala)

Synthesis of Ac-Glu-Ala-Ile-pTyr-Ala-Ala-Pro-Phe-PEG-NH_2 (2.12)[150]



The phosphorylated peptide **2.12** was synthesised manually at 25 °C. The appropriate Fmoc amino acid (0.08 mmol, 5 eq), HBTU (30 mg, 0.08 mmol, 5 eq) and HOBt (11 mg, 0.08 mmol, 5 eq) were dissolved in DMF (0.3 mL) followed by addition of DIPEA (26 µL, 0.15 mmol, 10 eq). The resulting solution was mixed for 10 s before being added to the resin **2.16** (loading 0.63 mmol/g, 24 mg, 0.015 mmol, 1 eq) pre-swollen in DMF. The mixture was shaken for 1 h. The completion of each coupling was verified by the Kaiser test or the chloranil test for Fmoc-Pro-OH (see section 4.1.2). If needed, an extended cycle was repeated until completion of the reaction. Fmoc deprotection was performed as described in section 4.1.2. A preformed protected phosphoamino acid Fmoc-Tyr(HPO_3Bzl)-OH (46 mg, 0.08 mmol, 5 eq) was coupled to the resin for 4 h in order to incorporate the phosphorylated tyrosine residue. After the last amino acid coupling, the resin was capped using acetic anhydride and pyridine (50/50, 0.4 mL) for 1.5 h. Finally, the phosphorylated peptide was cleaved from the resin using Reagent K (0.4 mL) for 4 h and precipitated with cold diethyl ether (5 mL). The residue was collected by filtration, dissolved in CH_3CN/H_2O, lyophilised and purified by reverse phase preparative HPLC to afford **2.12** as an off-white solid (10 mg, 59%).

HPLC (method 1): t_R = 5.2 min (100% purity, ELSD)

MALDI-TOF$^+$/MS: m/z (%) 1132.9 $[M+H]^+$ (100)

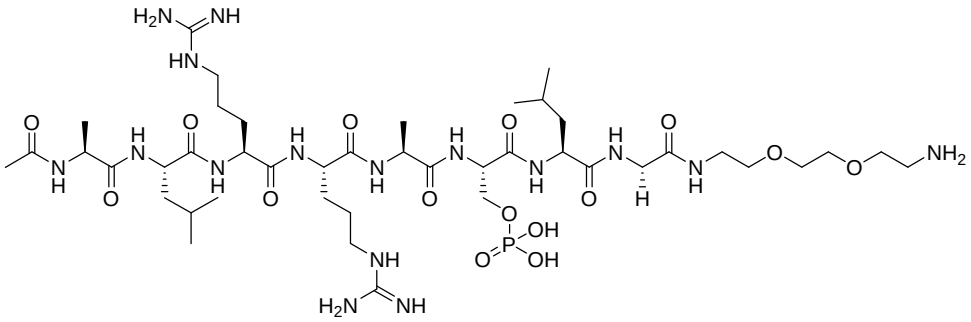
^{1}H NMR (800 MHz, D_2O): δ_H 7.35 (2H, t, *J* = 7.5 Hz, ArH-Phe), 7.32-7.20 (3H, m, ArH-Phe), 7.19-7.12 (2H, m, ArH-Tyr), 7.10-7.05 (2H, m, ArH-Tyr), 4.56 (1H, t, *J* = 7.7 Hz, H_α-Tyr), 4.51-4.43 (2H, m, H_α-Phe and H_α-Ala), 4.35 (1H, dd, *J* = 8.3, 5.7 Hz, H_α-Pro), 4.31-4.18 (3H, m, H_α-Glu and 2 x H_α-Ala), 4.13-3.96 (1H, m, H_α-Ile), 3.78 (2H, td, *J* = 10.2, 7.0 Hz, H_δ-Pro), 3.73-3.68 (2H, m, CH_2O), 3.66-3.53 (4H, m, CH_2O), 3.49-3.38 (2H, m, CH_2O), 3.38-3.19 (2H, m, C*H*$_2$NH), 3.19-3.13 (2H, m,

C*H*$_2$NH$_2$), 3.13-2.95 (4H, m, H$_\beta$-Phe and H$_\beta$-Tyr), 2.51-2.42 (2H, m, H$_\gamma$-Glu), 2.25-2.17 and 1.81-1.72 (2H, m, H$_\beta$-Pro), 2.10-1.89 (4H, m, H$_\beta$-Glu and H$_\gamma$-Pro), 2.00 (3H, s, CH$_3$-Ac), 1.81-1.72 (1H, m, H$_\beta$-Ile), 1.41-1.03 (2H, m, H$_\gamma$-Ile), 1.32 (3H, d, *J* = 7.1 Hz, CH$_3$-Ala), 1.31 (3H, d, *J* = 7.2 Hz, CH$_3$-Ala), 1.26 (3H, d, *J* = 7.1 Hz, CH$_3$-Ala), 0.91-0.71 (6H, m, CH$_3$-Ile)

^{13}C NMR (200 MHz, D$_2$O): δ_C 179.7 (COO-Glu), 177.3 (CO-Ala), 177.0 (CO-Ac), 176.5 (CO-Pro), 176.1 (CO-Ala), 176.0 (CO-Glu), 175.9 (CO-Ala), 175.8 (CO-Ile), 175.6 (CO-Phe), 174.8 (CO-Tyr), 153.7 and 153.7 (C-Tyr), 138.8 (C-Phe), 134.3 (C-Tyr), 133.0 (2 x ArH-Tyr), 131.9 (2 x ArH-Phe), 131.5 (2 x ArH-Phe), 129.9 (ArH-Phe), 123.2 (2 x ArH-Tyr), 72.2 (CH$_2$O), 72.1 (CH$_2$O), 71.3 (CH$_2$O), 69.1 (CH$_2$O), 63.1 (C$_\alpha$-Pro), 60.9 (C$_\alpha$-Ile), 57.9 (C$_\alpha$-Phe), 57.5 (C$_\alpha$-Tyr), 55.6 (C$_\alpha$-Glu), 52.3 (C$_\alpha$-Ala), 51.6 (C$_\alpha$-Ala), 50.5 (C$_\alpha$-Ala), 50.4 (C$_\delta$-Pro), 41.8 (CH$_2$NH$_2$), 41.5 (CH$_2$NH), 39.7 (C$_\beta$-Phe), 39.0 (C$_\beta$-Tyr), 38.7 (C$_\beta$-Ile), 32.6 (C$_\gamma$-Glu), 31.9 (C$_\beta$-Pro), 29.0 (C$_\beta$-Glu), 27.3 (C$_\gamma$-Pro), 27.2 (C$_\gamma$-Ile), 24.3 (CH$_3$-Ac), 19.5 (CH$_3$-Ala), 19.0 (CH$_3$-Ala), 18.0 (CH$_3$-Ala), 17.3 (CH$_3$-Ile), 12.7 (CH$_3$-Ile)

^{31}P NMR (146 MHz, D$_2$O): δ_P - 4.02

Synthesis of Ac-Ala-Leu-Arg-Arg-Ala-pSer-Leu-Gly-PEG-NH$_2$ (2.10)[150]



The phosphorylated peptide **2.10** was prepared as for peptide **2.12** except that 50 mg of resin **2.16** (experimental loading 0.63 mmol/g, 50 mg, 0.03 mmol, 1 eq) was used and a preformed protected phosphoamino acid Fmoc-Ser(HPO$_3$Bzl)-OH) (75 mg, 0.15 mmol, 5 eq) was coupled to the resin for 6 h. After purification by reverse phase preparative HPLC, the phosphorylated peptide **2.10** was obtained as a slight yellowish solid (12 mg, 38%).

HPLC (method 1): t_R = 3.4 min (100% purity, ELSD)

MALDI-TOF$^+$/MS: m/z (%) 1094.9 $[M+H]^+$ (100)

^{1}H NMR (800 MHz, D_2O): δ_H 4.56 (1H, t, *J* = 5.0 Hz, H_α-Ser), 4.38-4.26 (5H, m, 2 x H_α-Leu and H_α-Ala and 2 x H_α-Arg), 4.23 (1H, q, *J* = 7.2 Hz, H_α-Ala), 4.21-4.09 (2H, m, H_β-Ser), 3.91 (2H, s, H_α-Gly), 3.76 (2H, t, *J* = 5.1 Hz, CH_2O), 3.71 (4H, s, CH_2O), 3.64 (2H, t, *J* = 5.5 Hz, CH_2O), 3.45-3.39 (2H, m, C*H*$_2$NH), 3.25-3.17 (6H, m, 2 x H_δ-Arg and C*H*$_2$$NH_2$), 2.01 (3H, s, CH_3-Ac), 1.92-1.86 (2H, m, H_β-Arg), 1.86-1.80 (2H, m, H_β-Arg), 1.80-1.50 (10H, m, 2 x H_γ-Arg and 2 x H_β-Leu and 2 x H_γ-Leu), 1.41 (3H, d, *J* = 7.2 Hz, CH_3-Ala), 1.35 (3H, d, *J* = 7.2 Hz, CH_3-Ala), 0.94 (6H, dd, *J* = 10.1, 6.2 Hz, CH_3-Leu), 0.88 (6H, t, *J* = 5.8 Hz, CH_3-Leu)

^{13}C NMR (200 MHz, D_2O): δ_C 178.1 (CO-Ala), 177.9 (CO-Leu), 177.8 (CO-Leu), 177.3 (CO-Ala), 176.9 (CO-Ac), 176.2 (CO-Arg), 176.0 (CO-Arg), 174.1 (CO-Ser), 174.0 (CO-Gly), 159.4 (2 x C=NH-Arg), 72.3 (CH_2O), 72.2 (CH_2O), 71.5 (CH_2O), 69.1 (CH_2O), 66.6 and 66.6 (C_β-Ser), 57.0 and 56.9 (C_α-Ser), 55.9 (C_α-Arg), 55.8 (C_α-Arg), 55.5 (C_α-Leu), 55.0 (C_α-Leu), 52.8 (C_α-Ala), 52.5 (C_α-Ala), 45.3 (C_α-Gly), 43.2 (2 x C_δ-Arg), 42.2 (C_β-Leu), 42.2 (C_β-Leu), 41.8 (CH_2NH_2), 41.6 (CH_2NH), 30.8 (C_β-Arg), 30.7 (C_β-Arg), 27.1 (C_γ-Arg), 27.1 (C_γ-Arg), 27.0 (C_γ-Leu), 27.0 (C_γ-Leu), 24.9 (CH_3-Leu), 24.7 (CH_3-Leu), 24.2 (CH_3-Ac), 23.5 (CH_3-Leu), 23.3 (CH_3-Leu), 19.0 (CH_3-Ala), 19.0 (CH_3-Ala)

^{31}P NMR (146 MHz, D_2O): δ_P 0.29

4.2.2 Peptide printing

Ac-Glu-Ala-Ile-Tyr-Ala-Ala-Pro-Phe-PEG-NH_2 (**2.11**), Ac-Ala-Leu-Arg-Arg-Ala-Ser-Leu-Gly-PEG-NH_2 (**2.9**), Ac-Glu-Ala-Ile-pTyr-Ala-Ala-Pro-Phe-PEG-NH_2 (**2.12**), Ac-Ala-Leu-Arg-Arg-Ala-pSer-Leu-Gly-PEG-NH_2 (**2.10**) and Fluorescein-Glu-Ala-Ile-Tyr-Ala-Ala-Pro-Phe-PEG-NH_2 (**2.13**) were prepared as 5 mM, 1 mM, 0.5 mM, 0.1 mM, 50 µM, 10 µM, 1 µM and 100 nM solutions in printing buffer (50 mM sodium phosphate, pH 8.5). The peptides and phosphopeptides were printed, keeping humidity below 50%, onto CodeLink^{TM} activated glass slides using a Genetix QArrayMini contact printer, equipped with a 32-pin microarraying head. Each peptide and phosphopeptide was printed as 16 replicates (4 x 4 sub-arrays), with a typical spot size of 320 µm, a row pitch of 560 µm and a column pitch of 1120 µm. 20 aQu solid pins with a 150 µm diameter (K2783, Genetix) were used

and the following printing conditions were applied: 2 stamps per spot, 10 ms stamping time and 10 ms inking time. After printing, the slides were incubated, in the dark, at 25 °C for 16 h in a chamber containing a 3 M NaCl solution. Then, the slides were incubated in a pre-warmed blocking solution (0.1 M Tris, 50 mM ethanolamine, pH 9.0, 20 mL) at 50 °C for 30 min, rinsed with distilled water (2 x 20 mL), washed with 4x SSC/0.1% SDS (20 mL) pre-warmed to 50 °C for 30 min and rinsed with distilled water (20 mL) before being spin-dried by centrifugation (650 rpm for 10 min).

4.2.3 Antibody assays on peptide microarray

The slides were blocked with a pre-warmed buffer containing TPBS (0.05% Tween-20/PBS, pH 7.4)/1% BSA (20 mL) for 1 h at 30 °C. After washing in PBS (20 mL) for 5 min at 30 °C, slides, that were covered with a glass cover-slide and kept in a hybridisation chamber (Genomic Solutions), were probed with the primary antibody, monoclonal anti-phosphotyrosine clone PY-20 (mouse IgG2b isotype), diluted 1/500 (0.4 µg) in an antibody buffer containing TPBS/ 1% BSA (final volume of 200 µL) for 1 h at 30 °C. After washing with the antibody buffer (2 x 20 mL) for 10 min at 30 °C, a labeled secondary antibody, Cy3-goat anti-mouse IgG (H+L), diluted 1/1000 (0.1 µg) in the antibody buffer (final volume of 200 µL) was added and slides were incubated in the dark for 1 h at 30 °C. Slides were then washed in the antibody buffer (2 x 20 mL for 10 min at 30 °C) and rinsed with Tris buffer (20 mL) at 25 °C before being spin-dried by centrifugation (1000 rpm for 10 min).

4.2.4 Library synthesis

The 10,000-member PNA-encoded peptide library was synthesised by Laurent Bialy and Juan-José Díaz-Mochón.[158]

4.2.5 Solution-phase kinase assays[81]

Kinase reaction buffers:

Abl buffer (1x): 50 mM Tris-HCl (pH 7.5 at 25 °C), 10 mM $MgCl_2$, 1 mM EGTA, 2 mM DTT, 0.01% Brij 35.

Her2 and VEGFR2 buffer: 50 mM HEPES, pH 7.4, 3 mM $MgCl_2$, 3 mM $MnCl_2$, 1 mM DTT, 3 µM Na-orthovanadate.

The 10,000-member PNA-encoded library was incubated, in solution, at a final concentration of 50 µM and final volume of 100 µL, in a buffer containing 60 units of the selected kinase (Abl, Her2 or VEGFR2), 5 mM ATP, kinase reaction buffer and 0.1% BSA for 24 h at 30 °C.

Hybridisation protocol: before hybridisation, one volume of this buffer was diluted with one volume of GenHyb buffer (Genetix). Solutions were denatured by heating at 65 °C for 3 min; 100 µL of the solution was then added onto a 22,500 customised DNA chip from Oxford Gene Technology (Oxford, UK), which was covered with a glass cover-slide and kept in a hybridisation chamber (Genomic Solutions) at 60 °C. Temperature was then slowly lowered to 37 °C over a period of 20 h. Finally, the chips were washed twice (20 mL) for 10 min at 30 °C in the following buffer (100 mM Nacl, 10 mM citric acid, 0.7% (w/v) *N*-lauroylsarcosine sodium salt, 0.1 mM EGTA, pH 7.5), rinsed with distilled water (20 mL) and spin-dried by centrifugation (1000 rpm for 10 min). Slides were stored in the dark at 25 °C.

4.2.6 Antibody assays on the 22,500 customised DNA microarray

The assay was carried out in analogy to section 4.2.3 except that the labelled secondary antibody, Cy3-goat anti-mouse IgG (H+L) was diluted 1/500 (0.2 µg) in the antibody buffer (TPBS/ 1% BSA).

4.2.7 Stripping procedure[161]

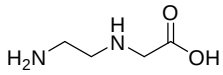
DNA chips could be stripped and re-hybridised up to three times. DNA chips were incubated in a stripping buffer [2.5 mM Na_2HPO_4, 0.1% (v/v) SDS, 20 mL] at 95 °C for 10 min. The procedure was repeated until complete probe removal. The DNA chips were then washed with distilled water (20 mL), rinsed with Tris buffer (20 mL)

and spin-dried by centrifugation (1000 rpm for 10 min). Before being re-hybridised and carrying out further experiments, the chip was scanned using both fluorescein and Cy3 filters.

4.3 Experimental for chapter 3

4.3.1 Synthesis of the Dde/Mmt-protected PNA monomers

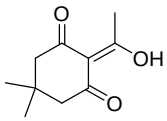
[(2-aminoethyl)amino]acetic acid (3.13)[4]



Chloroacetic acid (35.35 g, 374 mmol, 1 eq) was added portionwise at 0 °C to 1,2-ethanediamine (250 mL, 3.74 mol, 10 eq). The reaction was stirred at 25 °C for 16 h. Excess ethanediamine was evaporated *in vacuo* at 50 °C (higher temperatures led to decomposition) to give a yellow oil that was triturated with DMSO (400 mL). The resulting white solid was collected by filtration, washed with DMSO (3 x 50 mL), THF (3 x 50 mL) and diethyl ether (3 x 50 mL) before drying in a vacuum oven at 40 °C for 24 h to afford the named compound **3.13** as a white solid (29.1 g, 66%).

ESI$^+$/MS: m/z (%) 192.1 $[M+H]^+$ (28), 214.0 $[M+Na]^+$ (100)
^{1}H NMR (250 MHz, D_2O): δ_H 3.25 (2H, s, CH_2CO), 3.02 (2H, m, CH_2), 2.88 (2H, m, CH_2)
^{13}C NMR (62.5 MHz, D_2O): δ_C 178.2 (CO), 51.5 (CH_2), 46.5 (CH_2), 38.6 (CH_2)

2-(1-hydroxyethylidene)-5,5-dimethyl-1,3-cyclohexanedione (Dde-OH) (3.15)[216]

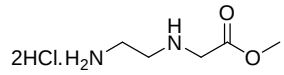


To a solution of dimedone (50 g, 357 mmol, 1 eq) in DMF (800 mL) were added acetic acid glacial (21 mL, 357 mmol, 1 eq), DCC (73.7 g, 357 mmol, 1 eq) and DMAP (4.4 g, 35.7 mmol, 0.1 eq). The reaction was stirred at 25 °C for 36 h. Precipitated dicyclohexylurea (DCU) was removed by filtration. After evaporation of

the solvent *in vacuo*, the residue was dissolved in EtOAc (300 mL), washed with 1 M aqueous $KHSO_4$ (3 x 100 mL), 10 % aqueous $NaHCO_3$ (3 x 100 mL) and brine (1 x 100 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo* to afford Dde-OH **3.15** as an orange oil (49.5 g, 76%).

HPLC (method 2): t_R = 3.3 min (100% purity, 254 nm)
ESI$^+$/MS: m/z (%) 183.2 $[M+H]^+$ (100)
^{1}H NMR (250 MHz, $CDCl_3$): δ_H 2.60 (3H, s, CH_3), 2.52 (2H, s, CH_2), 2.35 (2H, s, CH_2), 1.07 (6H, s, $C(CH_3)_2$)
^{13}C NMR (62.5 MHz, $CDCl_3$): δ_C 202.4 (C), 197.9 (CO), 195.2 (CO), 112.3 (C), 52.4 (CH_2), 46.9 (CH_2), 30.6 (C), 28.5 (CH_3), 28.2 (CH_3)

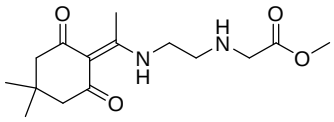
Methyl [(2-aminoethyl)amino]acetate dihydrochloride (3.14)[4]



To a stirred suspension of the acid **3.13** (28.2 g, 239 mmol, 1 eq) in MeOH (870 mL) at O °C was added dropwise thionyl chloride (87 mL, 1.20 mol, 5 eq). The reaction was then refluxed for 16 h before cooling to O °C. The resulting solid was filtered, washed with MeOH/DCM (1/1, 200 mL), diethyl ether (200 mL) and dried in a vacuum oven at 40 °C for 24 h. The named compound **3.14** was obtained as a white solid (43 g, 88%) and used without further purification.

HPLC (method 2): t_R = 0.9 min (100% purity, 254 nm)
ESI$^+$/MS: m/z (%) 133.1 $[M+H]^+$ (100)
^{1}H NMR (250 MHz, D_2O): δ_H 4.13 (2H, s, CH_2CO), 3.85 (3H, s, CH_3), 3.58-3.49 (2H, m, CH_2), 3.48-3.40 (2H, m, CH_2)
^{13}C NMR (62.5 MHz, D_2O): δ_C 167.8 (CO), 54.0 (CH_3), 48.0 (CH_2), 44.4 (CH_2), 35.7 (CH_2)

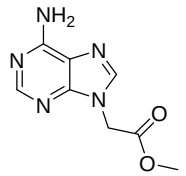
Methyl [(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl) amino]acetate (3.10)[4]



To a stirred solution of **3.14** (5 g, 24.4 mmol, 1 eq) and DIPEA (8.5 mL, 48.8 mmol, 2 eq) in DCM (80 mL) was added portionwise Dde-OH **3.15** (4.45 g, 24.4 mmol, 1 eq). The mixture was stirred at 25 °C for 16 h. After evaporation of the solvent *in vacuo*, the residue was dissolved in EtOAc (100 mL) and extracted with 1 M aqueous $KHSO_4$ (4 x 30 mL). The aqueous extracts were combined, brought to pH 9 with $NaHCO_3$ and extracted with DCM (4 x 30 mL). The organic phases were combined, washed with brine (1 x 30 mL), dried over $MgSO_4$ and concentrated *in vacuo* to afford the named compound **3.10** as an orange oil (4.6 g, 64%).

HPLC (method 2): t_R = 2.2 min (100% purity, 254 nm)
ESI$^+$/MS: m/z (%) 297.2 $[M+H]^+$ (100)
^{1}H NMR (250 MHz, $CDCl_3$): δ_H 13.47 (1H, br s, NH-Dde), 3.70 (3H, s, OCH_3), 3.50-3.40 (4H, m, $NH(CH_2)_2$), 2.91 (2H, t, *J* = 6.1 Hz, $NHCH_2CO$), 2.53 (3H, s, CCH_3), 2.33 (4H, s, CH_2-Dde), 1.79 (1H, br s, NH), 0.99 (6H, s, $(CH_3)_2$-Dde)
^{13}C NMR (62.5 MHz, $CDCl_3$): δ_C 197.1 (CO), 173.3 (C), 172.7 (CO), 107.9 (C), 52.7 (CH_2), 51.8 (CH_3), 50.2 (CH_2), 47.6 (CH_2), 43.3 (CH_2), 30.0 (C), 28.2 (CH_3), 18.1 (CH_3)

Methyl (6-amino-9*H*-purin-9-yl)acetate (3.17)[14]

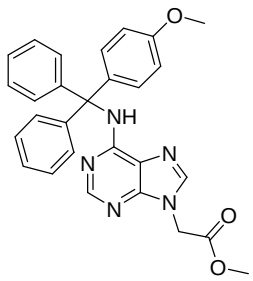


To a stirred suspension of adenine (50 g, 370 mmol, 1 eq) in DMF (800 mL) was added NaH (60 % dispersion in mineral oil, 13.7 g, 407 mmol, 1.1 eq). After stirring for 1 h at 25 °C, the mixture was cooled to 0 °C and methyl 2-bromoacetate (69 mL,

740 mmol, 2 eq) was added dropwise. The mixture was stirred at 25 °C for 16 h. After evaporation of the solvent *in vacuo*, the yellow oil was precipitated with H_2O, collected by filtration and washed with H_2O. The resulting solid was then dried in an oven at 40 °C for 16 h to afford the named compound **3.17** as a white solid (38 g, 50%).

HPLC (method 2): t_R = 0.9 min (100% purity, 254 nm)
ESI$^+$/MS: m/z (%) 208.2 $[M+H]^+$ (100)
^{1}H NMR (250 MHz, d_6-DMSO): δ_H 8.14 (1H, s, CH), 8.12 (1H, s, CH), 7.32 (2H, br s, NH_2), 5.09 (2H, s, CH_2), 3.70 (3H, s, CH_3)
^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 168.4 (CO), 155.9 (CNH_2), 152.6 (CH), 149.6 (C), 141.2 (CH), 118.2 (C), 52.4 (CH_3), 43.7 (CH_2)

Methyl (6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetate (3.18)[4]



To a stirred suspension of **3.17** (6 g, 29 mmol, 1 eq) in pyridine/DCM (1/1, 180 mL) was added NEM (3.7 mL, 29 mmol, 1 eq) and 4-monomethoxytrityl chloride (13.4 g, 43.5 mmol, 1.5 eq). The mixture was heated at 40 °C for 16 h. After evaporation of the solvents *in vacuo*, the residue was dissolved in EtOAc (360 mL), washed with 1 M aqueous $KHSO_4$ (90 mL), 10 % aqueous $NaHCO_3$ (90 mL), brine (90 mL) and dried over $MgSO_4$. After evaporation of the solvent *in vacuo*, the resulting orange oil was purified by silica gel chromatography using EtOAc as an eluent to afford the named compound **3.18** as a yellowish solid (11.4 g, 82%).

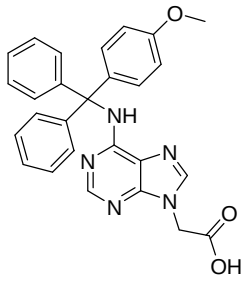
R_f (EtOAc) = 0.52
HPLC (method 2): t_R = 3.3 min (100% purity, 254 nm)

ESI$^+$/MS: m/z (%) 273.3 $[Mmt]^+$ (100), 480.5 $[M+H]^+$ (29)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 8.21 (1H, s, CH-pur), 7.91 (1H, s, CH-pur), 7.40-7.16 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 4.3 Hz, C*H*$COCH_3$-Mmt), 5.10 (2H, s, CH_2), 3.71 (3H, s, OCH_3-Mmt), 3.69 (3H, s, OCH_3)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 168.0 (CO), 157.3 (C), 153.1 (C), 151.0 (CH), 148.4 (C), 144.7 (C), 141.4 (CH), 136.7 (C), 129.5 (CH), 128.1 (CH), 127.3 (CH), 126.1 (CH), 119.5 (C), 112.6 (CH), 69.5 (C), 54.6 (CH_3), 52.1 (CH_3), 43.5 (CH_2)

(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetic acid (3.19)[4]



The ester **3.18** (4 g, 8.3 mmol, 1 eq) was suspended in 1 M aqueous NaOH (90 mL), refluxed for 2 h and stirred at 25 °C for 2 h. The solution was then cooled to 0 °C. Following the addition of 1 M aqueous $KHSO_4$ (110 mL), the mixture was stirred at 0 °C for 30 min. After extraction with EtOAc (2 x 100 mL), the organic phases were washed with brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo* to afford the named compound **3.19** as an off-white solid (3.3 g, 86%).

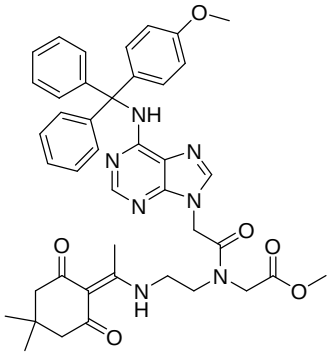
HPLC (method 2): t_R = 3.5 min (100% purity, 254 nm)

ESI$^-$/MS: m/z (%) 464.1 $[M-H]^-$ (100)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 8.20 (1H, s, CH-pur), 7.91 (1H, s, CH-pur), 7.38-7.16 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.9 Hz, C*H*$COCH_3$-Mmt), 4.97 (2H, s, CH_2) 3.71 (3H, s, OCH_3-Mmt)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 169.2 (CO), 157.6 (C), 153.4 (C), 151.2 (CH), 148.8 (C), 145.1 (C), 141.9 (CH), 137.0 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.5 (CH), 119.9 (C), 112.9 (CH), 69.8 (C), 54.9 (CH_3), 44.0 (CH_2)

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetyl]amino} acetate (3.20)[4]

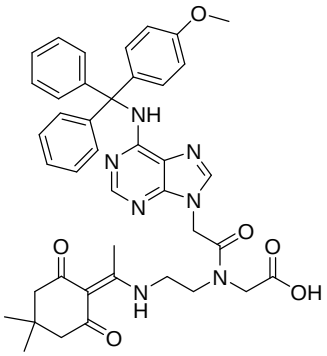


To a solution of **3.19** (3.48 g, 7.48 mmol, 1 eq) in CH_3CN (60 mL) was added HBTU (3.41 g, 8.98 mmol, 1.2 eq) and DIPEA (1.3 mL, 7.48 mmol, 1 eq). The mixture was stirred for 3 min and added to a solution of **3.10** (2.2 g, 7.48 mmol, 1 eq) in CH_3CN (60 mL). The resulting solution was stirred at 25 °C for 1 h and the solvent was removed *in vacuo*. The residue was dissolved in EtOAc (150 mL), washed with 1 M aqueous $KHSO_4$ (3 x 50 mL), 10% aqueous $NaHCO_3$ (50 mL), brine (50 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The residue was dissolved in a minimum amount of EtOAc, precipitated with hexanes, collected by filtration and dried *in vacuo*. The residue was then dissolved in a minimum amount of EtOAc, precipitated with an excess of petroleum ether, collected by filtration and dried *in vacuo* to afford the named compound **3.20** as a pale yellowish solid (4.75 g, 85%).

HPLC (method 2): t_R = 3.7 min (96% purity, 254 nm)
ESI$^+$/MS: m/z (%) 273.0 $[Mmt]^+$ (100), 744.3 $[M+H]^+$ (29)
^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: δ_H 13.26 and 13.16 (1H, br s, NH-Dde), 8.09 and 8.08 (1H, s, CH-pur), 7.88 (1H, s, CH-pur), 7.42-7.12 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.8 Hz, $CHCOCH_3$-Mmt), 5.25 and 5.08 (2H, s, CH_2COO), 4.52 and 4.12 (2H, s, CH_2CO), 3.86-3.76 (2H, m, CH_2N), 3.74 and 3.60 (3H, s, OCH_3), 3.72 (3H, s, OCH_3-Mmt), 3.57-3.46 (2H, m, CH_2N), 2.56 and 2.46 (3H, s, CH_3C), 2.29 and 2.26 (4H, s, CH_2-Dde), 0.94 and 0.93 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, d_6-DMSO) two rotamers: δ_C 196.5 and 196.4 (CO), 173.1 (CO), 169.8 (CO), 169.3 (CO), 167.7 and 167.0 (C), 157.6 (C), 153.3 (C), 151.1 (CH), 149.0 and 148.9 (C), 145.1 (C), 142.3 and 142.1 (CH), 137.0 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.4 (CH), 119.7 and 119.7 (C), 112.9 (CH), 107.3 and 107.2 (C), 69.8 (C), 54.9 (CH_3), 52.3 (CH_2), 51.8 (CH_3), 48.0 (CH_2), 46.5 (CH_2), 43.7 (CH_2), 41.0 (CH_2), 29.7 (C), 27.8 (CH_3), 17.4 and 17.1 (CH_3)

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetyl]amino}acetic acid (3.21)[4]



The ester **3.20** (3.56 g, 4.79 mmol, 1 eq) was suspended in a 2:1 (v/v) mixture of MeOH and 2 M Cs_2CO_3 (30 mL) and stirred for 1 h. After evaporation of MeOH *in vacuo*, the remaining solution was diluted with H_2O (150 mL), washed with EtOAc (50 mL), acidified with 1 M aqueous $KHSO_4$, extracted with DCM (3 x 100 mL). The organic extracts were washed with brine (100 mL), dried over $MgSO_4$ and the solvent was evaporated *in vacuo*. The residue was dissolved in a minimum amount of EtOAc, precipitated with hexanes, collected by filtration and dried *in vacuo* to afford the named compound **3.21** as a pale yellowish solid (2.53 g, 72%).

HPLC (method 2): t_R = 3.6 min (100% purity, 254 nm)

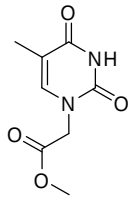
ESI$^-$/MS: m/z (%) 728.2 $[M-H]^-$ (100)

^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: δ_H 13.26 and 13.16 (1H, br s, NH-Dde), 8.10 and 8.08 (1H, s, CH-pur), 7.87 (1H, s, CH-pur), 7.38-7.12 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.8 Hz, $CHCOCH_3$-Mmt), 5.24 and 5.07 (2H, s, CH_2COO),

4.40 and 4.03 (2H, s, CH_2CO), 3.81-3.74 (2H, m, CH_2N), 3.71 (3H, s, OCH_3-Mmt), 3.60-3.47 (2H, m, CH_2N), 2.56 and 2.46 (3H, s, CH_3C), 2.29 and 2.26 (4H, s, CH_2-Dde), 0.99 and 0.93 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, d$_6$-DMSO) two rotamers: δ_C 196.5 (CO), 173.1 (CO), 170.7 (CO), 170.2 (CO), 167.6 and 166.8 (C), 157.6 (C), 153.3 (C), 151.1 (CH), 149.0 (C), 145.1 (C), 142.1 (CH), 137.0 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.5 (CH), 119.7 (C), 112.9 and 112.7 (CH), 107.3 and 107.2 (C), 69.8 (C), 54.9 (CH_3), 52.3 (CH_2), 49.1 (CH_2), 47.9 and 46.6 (CH_2), 43.7 (CH_2), 40.9 (CH_2), 29.7 (C), 27.8 and 27.5 (CH_3), 17.3 and 17.2 (CH_3)

Methyl (5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetate (3.23)[14]



To a stirred suspension of thymine (20 g, 158.6 mmol, 1 eq) and K_2CO_3 (21.9 g, 158.6 mmol, 1 eq) in DMF (500 mL) was added dropwise methyl 2-bromoacetate (19 mL, 206.2 mmol, 1.3 eq). The mixture was then stirred vigorously at 25 °C for 16 h, collected by filtration and evaporated to dryness *in vacuo*. The solid residue was cooled to 0 °C, treated with water (300 mL), 1 M aqueous $KHSO_4$ (20 mL) and stirred for 30 min. The precipitated was collected by filtration, washed with water (3 x 100 mL) to afford the named compound **3.23** as a white solid (18.16 g, 58%).

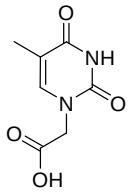
HPLC (method 2): t_R = 1.2 min (100% purity, 254 nm)

ESI$^+$/MS: m/z (%) 199.0 $[M+H]^+$ (100)

^{1}H NMR (250 MHz, d$_6$-DMSO): δ_H 11.41 (1H, br s, NH), 7.49 (1H, s, CH), 4.48 (2H, s, CH_2), 3.68 (3H, s, OCH_3), 1.75 (3H, s, CH_3)

^{13}C NMR (62.5 MHz, d$_6$-DMSO): δ_C 168.7 (CO), 164.3 (CO), 150.9 (CO), 141.5 (CH), 108.6 (C), 52.2 (OCH_3), 48.3 (CH_2), 11.9 (CH_3)

(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetic acid (3.24)[14]



The ester **3.23** (44 g, 0.22 mol, 1 eq) was suspended in 2 M aqueous NaOH (750 mL) and refluxed for 3 h. The solution was then cooled to 0 °C, treated with 1 M aqueous $KHSO_4$ (200 mL) and stirred for 30 min. The compound was collected by filtration, washed with water (3 x 150 mL) and freeze dried for 16 h to afford the named compound **3.24** as a white solid (20.35 g, 50%).

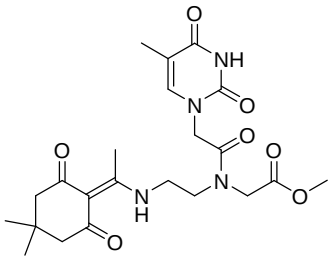
HPLC (method 2): t_R = 0.9 min (100% purity, 254 nm)

ESI$^-$/MS: m/z (%) 183.0 $[M-H]^-$ (100)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 13.12 (1H, br s, CO_2H), 11.36 (1H, br s, NH), 7.50 (1H, s, CH), 4.37 (2H, s, CH_2), 1.76 (3H, s, CH_3)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 169.6 (CO), 164.3 (CO), 150.9 (CO), 141.7 (CH), 108.3 (C), 48.3 (CH_2), 11.9 (CH_3)

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetyl]amino}acetate (3.25)[4]



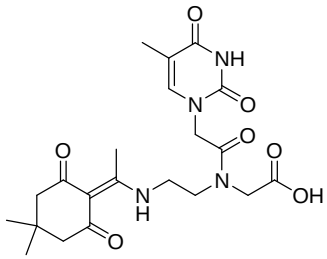
Compound **3.25** was synthesised from **3.24** (10.1 g, 54.9 mmol, 1 eq) in analogy to **3.20** and obtained as a buff coloured solid (19.6 g, 77%).

HPLC (method 2): t_R = 2.6 min (100% purity, 254 nm)

ESI$^+$/MS: m/z (%) 463.1 $[M+H]^+$ (100)

^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: δ_H 13.20 and 13.14 (1H, br s, NH-Dde), 11.31 and 11.30 (1H, br s, NH), 7.33 and 7.31 (1H, s, CH), 4.67 and 4.51 (2H, s, CH_2COO), 4.40 and 4.11 (2H, s, CH_2CO), 3.72 and 3.62 (3H, s, OCH_3), 3.71-3.65 (2H, m, CH_2N), 3.58-3.47 (2H, m, CH_2N), 2.53 and 2.47 (3H, s, CH_3C-Dde), 2.30 and 2.27 (4H, s, CH_2-Dde), 1.75 (3H, s, CH_3), 0.95 and 0.94 (6H, s, $(CH_3)_2$-Dde)
^{13}C NMR (62.5 MHz, CD_3COOD) two rotamers: δ_C 200.7 and 200.5 (CO), 176.7 and 176.5 (CO), 171.1 and 171.0 (CO), 170.2 and 169.5 (CO), 167.2 (C), 152.9 (CO), 143.9 (CH), 111.5 (C), 109.0 and 108.9 (C), 53.5 and 53.1 (CH_3), 52.7 (CH_2), 52.5 (CH_2), 50.9 (CH_2), 49.5 and 48.7 (CH_2), 42.6 and 42.0 (CH_2), 30.8 (C), 28.2 (CH_3), 18.9 and 18.7 (CH_3), 12.3 (CH_3)

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetyl]amino}acetic acid (3.26)[4]



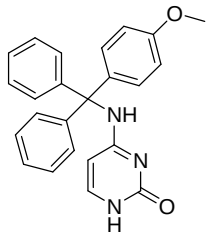
The ester **3.25** (5 g, 10.81 mmol, 1 eq) was suspended in 1:1 (v/v) mixture of MeOH and 2 M Cs_2CO_3 (100 mL) and stirred for 1 h. After evaporation of MeOH *in vacuo*, the aqueous phase was acidified with 4 M HCl (pH = 1) and the solvents were removed *in vacuo*. The residue was treated with hot 2-propanol and the hot suspension was passed through a filter. The filtrate was collected and 2-propanol was removed *in vacuo*. The residue was sonicated with water (25 mL), the resulting precipitated was collected by filtration and freeze dried to afford the named compound **3.26** as an off-white solid (1.94 g, 40%).

HPLC (method 2): t_R = 2.4 min (100% purity, 254 nm)
ESI$^-$/MS: m/z (%) 447.1 $[M-H]^-$ (100)
^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: δ_H 13.28-13.06 (1H, m, NH-Dde), 11.32 and 11.30 (1H, br s, NH), 7.34 and 7.32 (d, 1H, *J* = 0.9 Hz, CH), 4.66 and 4.50

(2H, s, CH_2COO), 4.28 and 4.02 (2H, s, CH_2CO), 3.75-3.61 (2H, m, CH_2N), 3.43-3.23 (2H, m, CH_2N), 2.53 and 2.47 (3H, s, CH_3C-Dde), 2.29 and 2.26 (4H, s, CH_2-Dde), 1.75 (3H, s, CH_3), 0.94 and 0.93 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, d_6-DMSO) two rotamers: δ_C 196.9 and 196.7 (CO), 173.5 (CO), 171.1 and 170.7 (CO), 168.4 and 167.7 (CO), 164.7 (C), 151.4 and 151.3 (CO), 142.4 and 142.3 (CH), 108.5 (C), 107.8 and 107.6 (C), 52.8 (2 x CH_2), 49.2 (CH_2), 48.1 and 46.8 (CH_2), 41.3 (CH_2), 30.1 (C), 28.2 (CH_3), 17.7 and 17.6 (CH_3), 12.4 and 12.3 (CH_3)

4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2(1*H*)-pyrimidinone (**3.28**)[19]



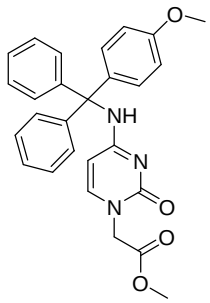
To a stirred suspension of cytosine (10 g, 90 mmol, 1 eq) in pyridine/DCM (1/1, 550 mL) was added NEM (11.5 mL, 90 mmol, 1 eq) and 4-monomethoxytrityl chloride (41.7 g, 135 mmol, 1.5 eq). The mixture was heated at 50 °C for 16 h. After evaporation of the solvents *in vacuo*, the resulting oil was precipitated with diethyl ether (200 mL). The resulting precipitated was collected by filtration, washed with water (3 x 100 mL), THF (3 x 100 mL) and dried *in vacuo* to afford the named compound **3.28** as a white solid (29 g, 84%).

HPLC (method 2): t_R = 3.1 min (100% purity, 254 nm)

ESI$^+$/MS: m/z (%) 273.0 $[Mmt]^+$ (100), 789.2 $[2M+Na]^+$ (22)

No NMR spectrum was recorded since the named compound **3.28** was insoluble in common deuterated solvents, as reported in the literature for Z-protected cytosine.[14]

Methyl (4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl) acetate (3.29)[4]



Compound **3.29** was synthesised from **3.28** (10 g, 26.1 mmol, 1 eq) in analogy to **3.18** and obtained as an off-white solid (10.4 g, 87%).

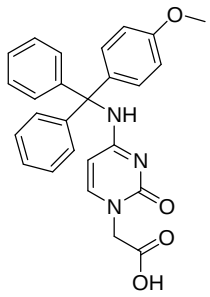
HPLC (method 2): t_R = 3.3 min (100% purity, 254 nm)

ESI$^-$/MS: m/z (%) 454.0 $[M-H]^-$ (100)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 8.40 (1H, br s, NH), 7.46 (1H, d, *J* = 6.8 Hz, CH-Cyt), 7.40-7.03 (12H, m, CH-Mmt), 6.84 (2H, d, *J* = 8.5 Hz, CH-Mmt), 6.20 (1H, d, *J* = 6.8 Hz, CH-Cyt), 4.33 (2H, s, CH_2), 3.72 (3H, s, OCH_3-Mmt), 3.61 (3H, s, OCH_3)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 169.1 (CO), 164.0 (C), 157.4 (C), 154.5 (CO), 148.8 (C and CH), 136.7 (C), 129.9 (CH), 128.5 (CH), 127.4 (CH), 126.1 (CH), 112.6 (CH), 95.8 (CH), 69.7 (C), 54.9 (CH_3), 51.9 (CH_3), 49.7 (CH_2)

(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl) acetic acid (3.30)[4]



The ester **3.29** (10.4 g, 22.8 mmol, 1 eq) was suspended in 2 M aqueous NaOH (100 mL) and the mixture was refluxed for 2 h. The solution was then cooled to 25 °C, washed with DCM (3 x 25 mL) and acidified with 2 M aqueous HCl until pH 2. The resulting precipitated was collected by filtration, washed with small portions of water until the filtrate was neutral (pH 7) and freeze dried for 16 h to afford the named compound **3.30** as a white solid (7.5 g, 74%).

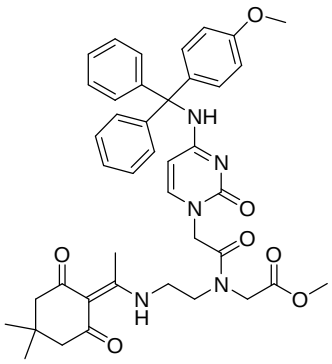
HPLC (method 2): t_R = 3.0 min (100% purity, 254 nm)

ESI⁻/MS: m/z (%) 440.1 $[M-H]^-$ (100)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 8.37 (1H, br s, NH), 7.46 (1H, d, *J* = 7.1 Hz, CH-Cyt), 7.36-7.05 (12H, m, CH-Mmt), 6.87 (2H, d, *J* = 8.1 Hz, CH-Mmt), 6.19 (1H, d, *J* = 7.0 Hz, CH-Cyt), 4.24 (2H, s, CH_2) 3.73 (3H, s, OCH_3-Mmt)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 170.4 (CO), 164.4 (C), 157.8 (C), 155.1 (CO), 148.4 (C), 145.3 (CH), 137.2 (C), 130.4 (CH), 129.0 (CH), 127.8 (CH), 126.6 (CH), 113.1 (CH), 96.0 (CH), 70.1 (C), 55.4 (CH_3), 50.1 (CH_2)

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl)acetyl] amino}acetate (3.31)[4]



To a solution of the acid **3.30** (2.3 g, 5.21 mmol, 1 eq) and PyBOP (2.98 g, 5.73 mmol, 1.1 eq) in DMF (15 mL) at 0 °C was added DIPEA (1.8 mL, 10.42 mmol, 2 eq) and then the Dde-protected backbone **3.10** (1.5 g, 5.21 mmol, 1 eq), the reaction was stirred under N_2 at 25 °C for 16 h. The solvent was removed *in vacuo*, the residue was re-dissolved in DCM (150 mL), washed with 1 M aqueous $KHSO_4$ (50 mL), 10 % aqueous $NaHCO_3$ (50 mL) and brine (50 mL). The organic phase was dried over $MgSO_4$ and the solvent was evaporated *in vacuo*. The residue was dissolved in a minimum amount of MeOH, precipitated with an excess of water, collected by filtration and freeze dried to give the named compound **3.31** (2.14 g, 57%) as a pale yellowish solid.

HPLC (method 2): t_R = 3.5 min (100% purity, 254 nm)

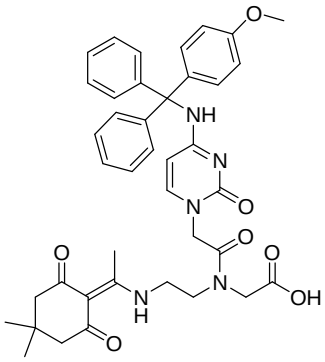
ESI$^+$/MS: m/z (%) 273.1 $[Mmt]^+$ (100), 720.3 $[M+H]^+$ (53)

^{1}H NMR (250 MHz, $CDCl_3$): δ_H 13.48 and 13.33 (1H, br s, NH-Dde), 7.39 and 6.97 (13H, m, 12 x CH-Mmt and CH-Cyt), 6.78 and 6.77 (2H, d, *J* = 8.8 Hz, $CHCOCH_3$-Mmt), 5.08 (1H, d, *J* = 7.1 Hz, CH-Cyt), 4.58 and 4.41 (2H, s, CH_2COO), 4.27 and 4.02 (2H, s, CH_2CO), 3.78-3.60 (7H, m, OCH_3-Mmt and 2 x CH_2N), 3.55 (3H, s, OCH_3), 2.54 and 2.50 (3H, s, CH_3C), 2.28 and 2.27 (4H, s, CH_2-Dde), 0.95 and 0.94 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, $CDCl_3$): δ_C 197.8 (CO), 174.1 and 173.9 (CNH), 169.9 and 169.8 (CO), 168.3 and 167.7 (CO), 165.5 (C), 158.7 (C), 155.9 (CO), 146.6 and

146.3 (CH), 143.8 (C), 135.5 (C), 129.9 (CH), 128.6 (CH), 128.3 (CH), 127.5 (CH), 113.6 (CH), 95.2 (CH), 70.8 (C), 55.2 (CH_3), 52.9 and 52.5 (CH_3), 52.7 (CH_2), 50.4 (CH_2), 49.0 and 48.6 (CH_2), 48.1 (CH_2), 41.3 and 40.7 (CH_2), 29.9 (C), 28.1 (CH_3), 17.8 and 17.7 (CH_3)

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl)acetyl] amino}acetic acid (3.32)[4]



Compound **3.32** was synthesised from **3.31** (3.2 g, 4.4 mmol, 1 eq) in analogy to **3.21** and obtained as a pale yellowish solid (2.4 g, 77%).

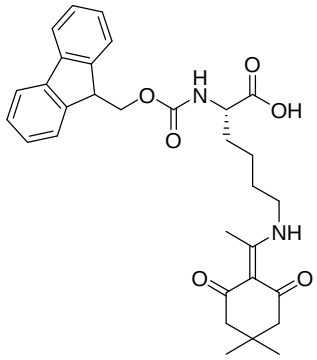
HPLC (method 2): t_R = 3.3 min (100% purity, 254 nm)

ESI⁻/MS: m/z (%) 704.2 $[M-H]^-$ (41)

^{1}H NMR (250 MHz, MeOH-d_4) two rotamers: δ_H 7.50-7.24 (11H, m, 10 x CH-Mmt and CH-Cyt), 7.21 (2H, d, *J* = 8.9 Hz, CH-Mmt), 6.97 (2H, d, *J* = 8.9 Hz, CH-Mmt), 5.35 and 5.31 (1H, d, *J* = 7.8 Hz, CH-Cyt), 4.88 and 4.75 (2H, s, CH_2COO), 4.34 and 4.14 (2H, s, CH_2CO), 3.85-3.64 (7H, m, OCH_3-Mmt and 2 x CH_2N), 2.57 and 2.55 (3H, s, CH_3C), 2.33 and 2.29 (4H, s, CH_2-Dde), 0.99 and 0.96 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, MeOH-d_4) two rotamers: δ_C 200.1 and 199.8 (CO), 176.0 (CNH), 172.1 (CO), 169.4 (CO), 164.9 (C), 161.3 and 161.1 (C), 152.7 (CO), 148.8 (CH), 143.7 (C), 135.0 (C), 131.3 (CH), 129.9 (CH), 129.8 (CH), 129.5 (CH), 115.2 (CH), 94.8 and 94.7 (CH), 74.4 (C), 55.9 (CH_3), 53.5 (CH_2), 50.7 (CH_2), 50.4 (CH_2), 49.9 (CH_2), 42.0 (CH_2), 31.1 (C), 28.5 (CH_3), 18.4 (CH_3)

(2*R*)-6-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}-2-{[(9*H*-fluoren-9-ylmethoxy)carbonyl]amino}hexanoic acid (3.38)[218]



To a stirred suspension of Fmoc-Lys-OH (5 g, 13.6 mmol, 1 eq) in ethanol (100 mL) was added Dde-OH **3.15** (5 g, 27.2 mmol, 2 eq) and TFA (104 µL, 1.36 mmol, 0.1 eq). The mixture was refluxed for 60 h and concentrated *in vacuo*. The orange residue was dissolved in EtOAc (200 mL), washed with 1 M aqueous $KHSO_4$ (2 x 100 mL), brine (1 x 100 mL), dried over $MgSO_4$ and the solvent was evaporated *in vacuo*. The resulting yellow oil was triturated with hexane to give a yellowish solid that was recrystallised from ethyl acetate/hexane to afford Fmoc-Lys(Dde)-OH **3.38** as an off-white solid (5 g, 69%).

HPLC (method 1): t_R = 9.3 min (100% purity, 254 nm)

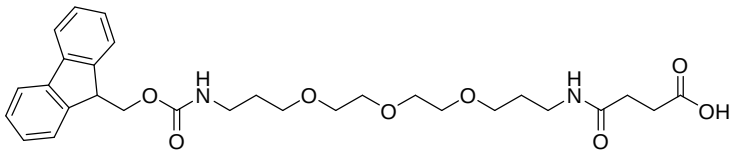
ESI$^+$/MS: m/z (%) 533.2 $[M+H]^+$ (100)

^{1}H NMR (250 MHz, d_6-DMSO): δ_H 13.34 (1H, br s, COOH), 12.72 (1H, br s, NH-Dde), 7.97 (2H, d, *J* = 7.4 Hz, CH-Fmoc), 7.81 (2H, d, *J* = 7.2 Hz, CH-Fmoc), 7.50 (2H, t, *J* = 7.4 Hz, CH-Fmoc), 7.42 (2H, d, *J* = 7.2 Hz, CH-Fmoc), 5.84 (1H, br s, NH), 4.41-4.38 (1H, m, CH), 4.33 (2H, d, *J* = 8.2 Hz, CH_2-Fmoc), 4.09-3.95 (1H, m, CH-Fmoc), 3.57-3.49 (2H, m, CH_2-NH), 2.55 (3H, s, CH_3-Dde), 2.32 (4H, s, CH_2-Dde), 1.91-1.72 (2H, m, CH_2-Lys), 1.71-1.56 (2H, m, CH_2-Lys), 1.55-1.37 (2H, m, CH_2-Lys), 1.00 (6H, s, $(CH_3)_2$-Dde)

^{13}C NMR (62.5 MHz, d_6-DMSO): δ_C 196.3 (CO-Dde), 173.8 (COOH), 172.7 (C=*C*NH), 156.1 (COO), 143.7 (C-Fmoc), 140.6 (C-Fmoc), 127.5 (CH-Fmoc), 127.0 (CH-Fmoc), 125.2 (CH-Fmoc), 120.0 (CH-Fmoc), 106.8 (*C*=CNH), 65.5 (CH_2-Fmoc), 53.6 (CH-Lys), 52.2 (2 x CH_2-Dde), 46.6 (CH-Fmoc), 42.4 (CH_2NH), 30.3

(CH_2-Lys), 29.6 (CH_2-Lys), 28.0 ($C(CH_3)_2$), 27.8 ($C(CH_3)_2$), 23.0 (CH_2-Lys), 17.3 (CH_3-Dde)

1-(9*H*-fluoren-9-yl)-3,19-dioxo-2,8,11,14-tetraoxa-4,18-diazadocosan-22-oic acid (Fmoc-PEG spacer-OH) (3.43)[219]



3-{2-[2-(3-aminopropoxy)ethoxy]ethoxy}-1-propanamine (2.2 g, 10 mmol, 1 eq) was dissolved in CH_3CN (50 mL) and succinic anhydride (1 g, 10 mmol, 1 eq) in CH_3CN (25 mL) was added dropwise over an hour. The reaction was allowed to proceed for an additional 3 h at 25 °C. After the waxy product settled, the organic solvent was decanted. The product was re-diluted in a mixture of H_2O and CH_3CN (1/1, 100 mL) and cooled to 0 °C for 30 min before Fmoc-OSu (4.4 g, 13 mmol, 1.3 eq) in CH_3CN (25 mL) was added. Enough DIPEA was then added to maintain the pH around 8-9 throughout the reaction. After stirring for 16 h at 25 °C, the solvents were removed *in vacuo*. The resultant product was dissolved in 10% aqueous $NaHCO_3$ (100 mL) and washed with EtOAc (25 mL). The aqueous phase was then acidified with 1 M HCl to pH 2 and extracted with EtOAc (3 x 50 mL). The combined organic phases were washed with H_2O (50 mL), dried over $MgSO_4$ and evaporated *in vacuo* to afford Fmoc-PEG spacer-OH **3.43** as a waxy product (4 g, 74%).

HPLC (method 2): t_R = 3.4 min (100% purity, 254 nm)

ESI$^-$/MS: m/z (%) 319.0 $[Fmoc]^-$ (100), 541.2 $[M-H]^-$ (87)

^{1}H NMR (250 MHz, MeOH-d_4): δ_H 7.79 (2H, d, *J* = 7.1 Hz, CH-Fmoc), 7.64 (2H, d, *J* = 7.1 Hz, CH-Fmoc), 7.39 (2H, t, *J* = 7.4 Hz, CH-Fmoc), 7.32 (2H, dd, *J* = 7.4, 1.2 Hz, CH-Fmoc), 4.36 (2H, d, *J* = 6.7 Hz, CH_2-Fmoc), 4.19 (1H, t, *J* = 6.6 Hz, CH-Fmoc), 3.66-3.39 (12H, m, CH_2O), 3.33-3.11 (4H, m, CH_2NH), 2.62-2.49 (2H, m, CH_2CO), 2.48-2.35 (2H, m, CH_2CO), 1.72 (4H, quint., *J* = 6.5 Hz, CH_2)

^{13}C NMR (62.5 MHz, MeOH-d_4): δ_C 176.2 (CO), 174.5 (CO), 158.9 (CO), 145.4 (C-Fmoc), 142.7 (C-Fmoc), 128.8 (CH-Fmoc), 128.2 (CH-Fmoc), 126.2 (CH-Fmoc), 121.0 (CH-Fmoc), 71.5 (CH_2O), 71.2 (CH_2O), 69.9 (CH_2O), 69.7 (CH_2O), 67.5 (CH_2-Fmoc), 54.9 (CH-Fmoc), 39.1 (CH_2NHCOO), 37.9 (CH_2NHCO), 31.7 (CH_2COOH), 30.8 (CH_2), 30.4 (2 x CH_2)

4.3.2 Synthesis of PNA-encoded positional scanning libraries

4.3.2.1 General experimental procedures

Dde deprotection[20]

$NH_2OH.HCl$ (6.25 g, 9 mmol, 1.3 eq) and imidazole (4.59 g, 6.75 mmol, 1 eq) were dissolved in 25 mL of NMP and the mixture was sonicated until complete dissolution. Just before reaction, 5 volumes of this solution were diluted with 1 volume of DMF. The solution was added to the resin and the resulting mixture was shaken for 1 h at 25 °C. The resin was finally washed with DMF (3 x) and DCM (3 x).

Dde/Mmt-protected PNA monomer couplings[4]

Dde/Mmt-protected PNA monomer (5 eq) and PyBOP (4.8 eq) were dissolved in DMF (0.1 M) followed by the addition of NEM (10 eq). The resulting solution was mixed for 10 sec. before adding to the PEGA resin (1 eq) pre-swelled in DMF and the mixture was shaken for 6 h. Resins were washed with DMF (3 x), DCM (3 x), MeOH (3 x) and DCM (3 x).

This procedure was also used for the couplings of Fmoc-Rink amide linker, Fmoc-PEG spacer-OH, rhodamine and 5(6)-carboxyfluorescein.

Fmoc-amino acid couplings

Fmoc-amino acids (5 eq), PyBOP (4.8 eq) and HOBt (5 eq) were dissolved in DMF (0.1 M) followed by the addition of DIPEA (10 eq). The resulting solution was mixed for 10 sec. before adding to the PEGA resin (1 eq) pre-swelled in DMF and the mixture was shaken for 6 h. Resins were washed with DMF (3 x), DCM (3 x), MeOH (3 x) and DCM (3 x).

Aloc deprotection[222]

- **625-member PNA-encoded positional scanning library (first sub-library)**

The resin (8.25 µmol, 1 eq) was pre-swelled in dry DCM under N_2 for 20 min and washed with dry DCM (3 x 2 mL) under N_2. A solution of $PhSiH_3$ (25 µL, 0.198 mmol, 24 eq) in dry DCM (350 µL) was then added and the resin was stirred manually under N_2. A solution of $Pd(PPh_3)_4$ (0.95 mg, 0.825 µmol, 0.1 eq) in dry DCM (1 mL) was added and the resin was shaken for 15 min with a balloon of N_2. The resin was washed with dry DCM under N_2 (3 x 2 mL) and the process was repeated once.

- **50,625-member PNA-encoded positional scanning library**

The same protocol was applied for each sub-library but using 0.116 mmol of resin (1 eq).

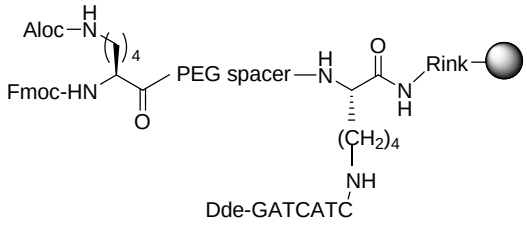
Cleavage of a sub-library from the resin

Each sub-library was cleaved from the resin using a TFA/TIS/DCM (90/5/5) solution (1 mL or 6 mL for the 625 or 50,625-member library respectively) for 2 h and precipitated with cold diethyl ether. The residue was collected by filtration and dried *in vacuo* for 2 h to afford the expected compound as a solid.

4.3.2.2 Library synthesis

- **625-member PNA-encoded positional scanning library (first sub-library)**

Synthesis of the common part of the library (3.45)



Deprotection and coupling protocols are described in section 4.3.2.1. Fmoc-Rink-amide linker (96 mg, 0.178 mmol, 5 eq) was first attached to PEGA resin (89 mg, 0.0356 mmol, 1 eq) following stepwise couplings with Fmoc-Lys(Dde)-OH **3.38** (95 mg, 0.178 mmol, 5 eq), Fmoc-PEG spacer-OH **3.43** (97 mg, 0.178 mmol, 5 eq) and

Fmoc-Lys(Aloc)-OH (81 mg, 0.178 mmol, 5 eq). Then, the Dde group was removed and seven common PNA monomers GATCATC (from N → C) (0.178 mmol, 5 eq) were coupled to the resin to afford the named compound **3.45**.

Loading: determined by a quantitative Fmoc test, 0.12 mmol/g (theoretical loading: 0.18 mmol/g).

Characterisation: a small sample of resin **3.45** was cleaved using a TFA/TIS/DCM (90/5/5) solution (500 µL) for 1 h.

HPLC (method 4): t_R = 6.7 min (91% purity, ELSD)

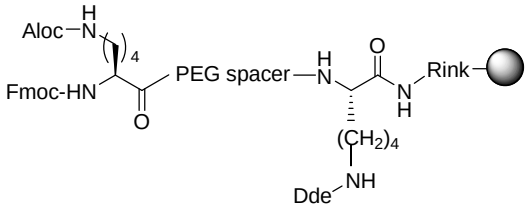
MALDI-TOF$^+$/MS: m/z (%) 2923.1 $[M+H]^+$ (100)

Synthesis of the first sub-library (3.51)

Deprotection, coupling and cleavage protocols are described in section 4.3.2.1. The synthesis was then carried out as follows: [1] split the common part of the library **3.45** into five different pools (6.6 µmol of resin **3.45** per pool); [2] Fmoc deprotection (see section 4.1.2); [3] Fmoc-amino acid couplings (33 µmol per pool, 5 eq); [4] Take one fourth of the amount of resin in each pool; [5] Dde deprotection; [6] Dde/Mmt-protected PNA monomer couplings (8.25 µmol per pool, 5 eq); repeat steps [5] and [6] four times; [7] the five resin pools were then mixed; repeat steps from [1] to [3] followed by step [7] three times. Then, Fmoc-γ-Abu-OH (13.5 mg, 41.25 µmol, 5 eq) and rhodamine (20 mg, 41.25 µmol, 5 eq) were respectively coupled followed by Aloc-deprotection and coupling with 5(6)-carboxyfluorescein (16 mg, 41.25 µmol, 5 eq). Before cleavage, the sub-library was treated with 20% piperidine in DMF (500 µL, 2 cycles of 10 min),[151] washed with DMF (2 x 500 µL) and DCM (2 x 500 µL). Final cleavage from the resin was carried out using a TFA/TIS/DCM (90/5/5) solution (1 mL) for 2 h to afford the first sub-library **3.51** as purple solid (28.3 mg, 67%).

- **50,625-member PNA-encoded positional scanning library**

Synthesis of the common part of the library (3.44)



Deprotection and coupling protocols are described in section 4.3.2.1. Fmoc-Rink-amide linker (2.7 g, 5 mmol, 5 eq) was first attached to PEGA resin (2.52 g, 1 mmol, 1 eq) following stepwise couplings with Fmoc-Lys(Dde)-OH **3.38** (2.66 g, 5 mmol, 5 eq), Fmoc-PEG spacer-OH **3.43** (2.7 g, 5 mmol, 5 eq) and Fmoc-Lys(Aloc)-OH (2.26 g, 5 mmol, 5 eq) to afford the named compound **3.44**.

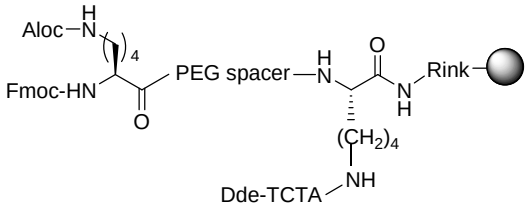
Loading: determined by a quantitative Fmoc test, 0.22 mmol/g (theoretical loading: 0.28 mmol/g).

Characterisation: a small sample of resin **3.44** was cleaved using a TFA/TIS/DCM (90/5/5) solution (500 µL) for 1 h.

HPLC (method 2): t_R = 3.3 min (86% purity, 254 nm)

MALDI-TOF$^+$/MS: m/z (%) 1046.2 $[M+H]^+$ (100)

Synthesis of the common part of the first sub-library (3.52)



Common part of the library **3.44** (0.529 g, 0.116 mmol, 1 eq) was washed with DMF (2 x 5 mL), DCM (6 x 5 mL) and pre-swelled in DCM for 30 min. The Dde group was removed and four common PNA monomers (TCTA, from N → C) (0.58 mmol, 5 eq) were coupled to the resin (see section 4.3.2.1) to afford the named compound **3.52**.

Characterisation: a small sample of resin **3.52** was cleaved using a TFA/TIS/DCM (90/5/5) solution (500 µL) for 1 h.

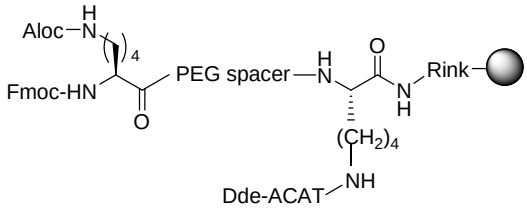
HPLC (method 2): t_R = 2.9 min (78% purity, 254 nm)

MALDI-TOF$^+$/MS: m/z (%) 2104.6 $[M+H]^+$ (100)

Synthesis of the first sub-library (3.56)

Deprotection, coupling and cleavage protocols are described in section 4.3.2.1. Synthesis of the first sub-library was carried out as follows: [1] split the common part of the first sub-library **3.52** into fifteen different pools (7.7 µmol of resin **3.52** per pool); [2] Fmoc deprotection (see section 4.1.2); [3] Fmoc-amino acid couplings (38.5 µmol per pool, 5 eq); Amino acids in the first position were then encoded by eight PNA monomers (Table 3.2). [4] Dde deprotection; [5] Dde/Mmt-protected PNA monomer couplings (38.5 µmol, 5 eq); repeat steps [4] and [5] seven times; [6] the fifteen resin pools were then mixed; repeat steps from [1] to [3] followed by step [6] three times. Then, Fmoc-γ-Abu-OH (189 mg, 0.58 mmol, 5 eq) and rhodamine (283 mg, 0.58 mmol, 5 eq) were respectively coupled followed by Aloc-deprotection and coupling with 5(6)-carboxyfluorescein (218 mg, 0.58 mmol, 5 eq). Before cleavage, the sub-library was treated with 20% piperidine in DMF (6 mL, 2 cycles of 10 min),[151] washed with DMF (2 x 6 mL) and DCM (2 x 6 mL). Final cleavage from the resin was carried out using a TFA/TIS/DCM (90/5/5) solution (6 mL) for 2 h to afford the first sub-library **3.56** as a magenta solid (176 mg, 30%).

Synthesis of the common part of the second sub-library (3.53)



The named compound **3.53** was synthesised in analogy to **3.52** but using the four common PNA monomers (ACAT, from N → C).

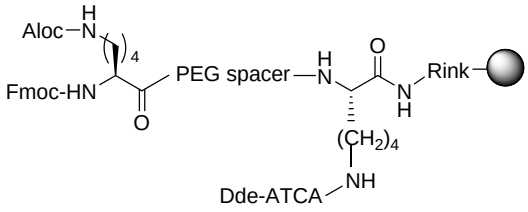
HPLC (method 2): t_R = 2.9 min (84% purity, 254 nm)

MALDI-TOF$^+$/MS: m/z (%) 2113.9 $[M+H]^+$ (100)

Synthesis of the second sub-library (3.57)

The procedure applied to **3.56** was followed except that amino acids in the second position of the tetrapeptide were encoded by eight PNA monomers (Table 3.2) to afford the second sub-library **3.57** as a magenta solid (175 mg, 29%).

Synthesis of the common part of the third sub-library (3.54)



The named compound **3.54** was synthesised in analogy to **3.52** but using the four common PNA monomers (ATCA, from N → C).

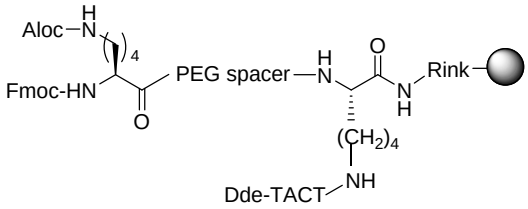
HPLC (method 2): t_R = 2.9 min (83% purity, 254 nm)

MALDI-TOF$^+$/MS: m/z (%) 2114.5 $[M+H]^+$ (100), 2136.5 $[M+Na]^+$ (18)

Synthesis of the third sub-library (3.58)

The procedure applied to **3.56** was followed except that amino acids in the third position of the tetrapeptide were encoded by eight PNA monomers (Table 3.2) to afford the third sub-library **3.58** as a magenta solid (256 mg, 43%).

Synthesis of the common part of the fourth sub-library (3.55)



The named compound **3.55** was synthesised in analogy to **3.52** but using the four common PNA monomers (TACT, from N → C).

HPLC (method 2): t_R = 3.0 min (85% purity, 254 nm)

MALDI-TOF$^+$/MS: m/z (%) 2103.9 $[M+H]^+$ (100), 2125.9 $[M+Na]^+$ (75)

Synthesis of the fourth sub-library (3.59)

The procedure applied to **3.56** was followed except that amino acids in the fourth position of the tetrapeptide were encoded by eight PNA monomers (Table 3.2) to afford the fourth sub-library **3.59** as a magenta solid (206 mg, 35%).

4.3.3 Enzymatic assays

DNA microarray production

3'-amino modified 12-mer DNAs oligomers were purchased from Sigma-Genosys (UK) (Table 3.3) or Bioneer (Korea) (Table 3.4). The dried oligonucleotides were centrifuged and re-suspended in 1 x TE (10 mM Tris, pH 7.5-8.0, 1mM EDTA) to give a 200 µM final concentration. Before printing, 10 µL of the solution of oligomers and 10 µL of spotting solution (2 M NaCl, 3 M $NaHCO_3$, pH 8) were mixed onto a 384-well plate. DNA printing was performed using a Genetix QArrayMini (Genetix) contact printer. One solid pin (20 aQu solid pins with a 150 µm diameter, Genetix) was used in order to print 3'-amino modified DNA oligomers (18 mers) on aldehyde-coated glass slides keeping humidity at 70 %. For the 625 member library, each DNA oligomer was printed in five replicates (1 x 5 sub-arrays), and this pattern was duplicated across the microarray. A different pattern was used for the 50,625 member PNA-encoded positional scanning library (see below). Then, the microarrays were kept at 25°C for 16 h in a chamber containing a filter paper saturated with a 3 M NaCl solution and stored in the dark. Finally, the chips were washed with 0.2 % Sodium Dodecyl Sulphate (SDS) (2 x 2 min) and distilled water (2 x 2 min). Imine reduction was performed by washing the slides for 5 min with a solution containing 50 mg of $NaBH_4$ in 15 mL of PBS and 5 mL of EtOH. The chips were then washed with distilled water (1 min), 0.2 % SDS (3 x 1 min) and distilled water (1 min). After spin-drying by centrifugation (4000 rpm for 6 min), the chips were stored in the dark at 25°C before hybridisation was carried out.

Arraying pattern used for the 50,625 member PNA-encoded positional scanning library

Each DNA oligomer was printed in five replicates (1 x 5 sub-arrays). A typical spot size of 200 µm, a row pitch of 670 µm and a column pitch of 670 µm were used as

well as the following printing conditions: 1 stamp per spot, 10 ms stamping time and 10 ms inking time.

Hybridisation onto DNA microarrays

The procedure described in section 4.2.5 was adapted. The hybridisation temperature was slowly lowered to 30 °C over a period of 5 h or 9 h for the 625 and 50,625-member PNA-encoded positional scanning libraries respectively.

Enzyme buffers

- *trypsin*: 67 mM sodium phosphate buffer, pH 7.6 at 25°C (625-member PNA-encoded library **3.51**) or 0.9% sodium chloride aqueous solution (50,625-member PNA-encoded library, sub-libraries **3.56**, **3.57**, **3.58** and **3.59**).
- *subtilisin A*: 10 mM sodium acetate buffer, 5 mM calcium acetate, pH 7.5.
- *chymopapain*: 50 mM sodium acetate, pH 6.2, 2 mM cysteine, 0.1 mM EDTA.
- *pepsin*: 0.2 M sodium citrate buffer (pH 4.5).

General procedure for fluorometric assays

The proteases subtilisin A, trypsin (TPCK treated, from bovine pancreas), chymopapain (from papaya latex) and pepsin (from porcine gastric mucosa) were purchased from Sigma.

The 625-member PNA encoded first sub-library **3.51** or one pool of the 50,625-member PNA-encoded positional scanning library (sub-libraries **3.56**, **3.57**, **3.58** or **3.59**) was diluted in the corresponding protease buffer (see section 4.3.3) (final volume of 125 µL) to give a final concentration of 100 or 50 µM for the 625 and 50,625-member libraries respectively. The enzymatic reaction was started by addition of the protease (60 µL) at a final concentration of 10 µM or 20 nM for the 625 and 50,625-member libraries respectively. The solution was agitated for 30 sec. and the fluorescence of the solution was measured by spectrofluorometry using an excitation wavelength of 480 nm and an emission wavelength of 520 nm. The change of fluorescence intensity was followed over 30 min or 1 h. Experiments were run in duplicate.

General procedure for proteolytic assays

The 625-member PNA encoded first sub-library **3.51** (solubilised in distilled water) was diluted in the corresponding protease buffer (see section 4.3.3) to give a final concentration of 100 µM and a final volume of 200 µL. The enzymatic reaction was started by addition of the protease (20 µL) at a final concentration of 10 µM. The solution was then incubated at 37 °C for 2 h and diluted with distilled water to give a final concentration of 10 µM. One volume of this solution was diluted with one volume of the GenHyb hybridisation buffer (see section 4.2.5) and used for hybridisation on DNA microarrays (see section 4.3.3).

The above protocol was adapted for the 50,625-member PNA encoded positional scanning library. Mother solutions of the sub-libraries **3.56**, **3.57**, **3.58** and **3.59** were first diluted 10 times in DMSO (up to 10%) and distilled water, and then in the corresponding protease buffer (see section 4.3.3) to give a final concentration of 50 µM and a final volume of 100 µL. A 20 nM protease final concentration was added and the solution was incubated at 37 °C for 2 h. Protease concentrations were calculated by UV spectroscopy using λ_{max} = 280 nm and the corresponding protease $E^{1\%}$.

REFERENCES

1. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991,** *254*, 1497-1500.
2. Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W. *Angew. Chem., Int. Ed.* **1998,** *37*, 2796-2823.
3. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993,** *365*, 566-568.
4. Bialy, L.; Diaz-Mochon, J. J.; Specker, E.; Keinicke, L.; Bradley, M. *Tetrahedron* **2005,** *61*, 8295-8305.
5. Nielsen, P. E.; Egholm, M. *Current Issues Molec. Biol.* **1999,** *1*, 89-104.
6. Wittung, P.; Nielsen, P. E.; Norden, B. *Biochemistry* **1997,** *36*, 7973-7979.
7. Demidov, V. V.; Yavnilovich, M. V.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92*, 2637-2641.
8. Egholm, M.; Christensen, L.; Dueholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E. *Nucleic Acids Res.* **1995,** *23*, 217-222.
9. Griffith, M. C.; Risen, L. M.; Greig, M. J.; Lesnik, E. A.; Sprankle, K. G.; Griffey, R. H.; Kiely, J. S.; Freier, S. M. *J. Am. Chem. Soc.* **1995,** *117*, 831-832.
10. Veselkov, A. G.; Demidov, V. V.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Nature* **1996,** *379*, 214.
11. Kuhn, H.; Hu, Y.; Frank-Kamenetskii, M. D.; Demidov, V. V. *Biochemistry* **2003,** *42*, 4985-4992.
12. Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchard, O.; Sonnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994,** *48*, 1310-1313.
13. Casale, R.; Jensen, I. S.; Egholm, M. *Peptide Nucleic Acids (2nd Edition); Horizon Bioscience* **2004**, 61-76.
14. Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. *J. Org. Chem.* **1994,** *59*, 5767-5773.
15. Pothukanuri, S.; Pianowski, Z.; Winssinger, N. *Eur. J. Org. Chem.* **2008**, 3141-3148.
16. Wojciechowski, F.; Hudson, R. H. E. *J. Org. Chem.* **2008,** *73*, 3807-3816.
17. Porcheddu, A.; Giacomelli, G.; Piredda, I.; Carta, M.; Nieddu, G. *Eur. J. Org. Chem.* **2008**, 5786-5797.
18. Thomson, S. A.; Josey, J. A.; Cadilla, R.; Gaul, M. D.; Hassman, C. F.; Luzzio, M. J.; Pipe, A. J.; Reed, K. L.; Ricca, D. J.; Wiethe, R. W.; Noble, S. A. *Tetrahedron* **1995,** *51*, 6179-6194.
19. Breipohl, G.; Knolle, J.; Langner, D.; O'Malley, G.; Uhlmann, E. *Bioorg. Med. Chem. Lett.* **1996,** *6*, 665-670.
20. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M. *Org. Lett.* **2004,** *6*, 1127-1129.

21. Bycroft, B. W.; Chan, W. C.; Chhabra, S. R.; Teesdale-Spittle, P. H.; Hardy, P. M. *J. Chem. Soc., Chem. Commun.* **1993**, 776-777.
22. Ray, A.; Norden, B. *FASEB J.* **2000,** *14*, 1041-1060.
23. Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G. *Science* **1992,** *258*, 1481-1485.
24. Tripathi, S.; Chaubey, B.; Barton, B. E.; Pandey, V. N. *Virology* **2007,** *363*, 91-103.
25. Tan, X.-X.; Actor, J. K.; Chen, Y. *Antimicrob. Agents Chemother.* **2005,** *49*, 3203-3207.
26. Boffa, L. C.; Scarfi, S.; Mariani, M. R.; Damonte, G.; Allfrey, V. G.; Benatti, U.; Morris, P. L. *Cancer Res.* **2000,** *60*, 2258-2262.
27. Larsen, H. J.; Nielsen, P. E. *Nucleic Acids Res.* **1996,** *24*, 458-463.
28. Janowski, B. A.; Kaihatsu, K.; Huffman, K. E.; Schwartz, J. C.; Ram, R.; Hardy, D.; Mendelson, C. R.; Corey, D. R. *Nat. Chem. Biol.* **2005,** *1*, 210-215.
29. Liebling, M. R.; Jou, N.-T.; Fang, W.; Louie, J. S. *Mol. Biotechnol.* **2003,** *25*, 229-240.
30. Diviacco, S.; Rapozzi, V.; Xodo, L.; Helene, C.; Quadrifoglio, F.; Giovannangeli, C. *FASEB J.* **2001,** *15*, 2660-2668.
31. Boutimah-Hamoudi, F.; Leforestier, E.; Senamaud-Beaufort, C.; Nielsen, P. E.; Giovannangeli, C.; Saison-Behmoaras, T. E. *Nucleic Acids Res.* **2007,** *35*, 3907-3917.
32. Fabani, M. M.; Gait, M. J. *RNA* **2008,** *14*, 336-346.
33. Koppelhus, U.; Nielsen, P. E. *Adv. Drug Deliv. Rev.* **2003,** *55*, 267-280.
34. Shammas, M. A.; Simmons, C. G.; Corey, D. R.; Shmookler Reis, R. J. *Oncogene* **1999,** *18*, 6191-6200.
35. Hamilton, S. E.; Simmons, C. G.; Kathiriya, I. S.; Corey, D. R. *Chem. Biol.* **1999,** *6*, 343-351.
36. Diaz-Mochon, J. J.; Bialy, L.; Watson, J.; Sanchez-Martin, R. M.; Bradley, M. *Chem. Commun.* **2005**, 3316-3318.
37. Koppelhus, U.; Awasthi, S. K.; Zachar, V.; Holst, H. U.; Ebbesen, P.; Nielsen, P. E. *Antisense Nucleic Acid Drug Dev.* **2002,** *12*, 51-63.
38. Rebuffat, A. G.; Nawrocki, A. R.; Nielsen, P. E.; Bernasconi, A. G.; Bernal-Mendez, E.; Frey, B. M.; Frey, F. J. *FASEB J.* **2002,** *16*, 1426-1428.
39. Dragulescu-Andrasi, A.; Rapireddy, S.; He, G.; Bhattacharya, B.; Hyldig-Nielsen, J. J.; Zon, G.; Ly, D. H. *J. Am. Chem. Soc.* **2006,** *128*, 16104-16112.
40. Mehiri, M.; Upert, G.; Tripathi, S.; Di Giorgio, A.; Condom, R.; Pandey, V. N.; Patino, N. *Oligonucleotides* **2008,** *18*, 245-256.
41. Orum, H.; Nielsen, P. E.; Jorgensen, M.; Larsson, C.; Stanley, C.; Koch, T. *Biotechniques* **1995,** *19*, 472-480.
42. Demers, D. B.; Curry, E. T.; Egholm, M.; Sozer, A. C. *Nucleic Acids Res.* **1995,** *23*, 3050-3055.
43. Perry-O'Keefe, H.; Yao, X.-W.; Coull, J. M.; Fuchs, M.; Egholm, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93*, 14670-14675.
44. Stender, H. *Expert Rev. Mol. Diagn.* **2003,** *3*, 649-655.

45. Lansdorp, P. M.; Verwoerd, N. P.; van de Rijke, F. M.; Dragowska, V.; Little, M.-T.; Dirks, R. W.; Raap, A. K.; Tanke, H. J. *Hum. Mol. Genet.* **1996,** *5*, 685-691.
46. Igloi, G. L. *Expert Rev. Mol. Diagn.* **2003,** *3*, 17-26.
47. Thiede, C.; Bayerdoerffer, E.; Blasczyk, R.; Wittig, B.; Neubauer, A. *Nucleic Acids Res.* **1996,** *24*, 983-984.
48. Seitz, O. *Angew. Chem., Int. Ed.* **2000,** *39*, 3249-3252.
49. Socher, E.; Bethge, L.; Knoll, A.; Jungnick, N.; Herrmann, A.; Seitz, O. *Angew. Chem., Int. Ed.* **2008,** *47*, 9555-9559.
50. Merrifield, R. B. *J. Am. Chem. Soc.* **1963,** *85*, 2149-2154.
51. Geysen, H. M.; Meloen, R. H.; Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984,** *81*, 3998-4002.
52. Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985,** *82*, 5131-5135.
53. Nicolaou, K. C.; Xiao, X.-Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. *Angew. Chem., Int. Ed.* **1995,** *34*, 2289-2991.
54. Nicolaou, K. C.; Pfefferkorn, J. A.; Mitchell, H. J.; Roecker, A. J.; Barluenga, S.; Cao, G. Q.; Affleck, R. L.; Lillig, J. E. *J. Am. Chem. Soc.* **2000,** *122*, 9954-9967.
55. Frank, R. *Tetrahedron* **1992,** *48*, 9217-9232.
56. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. Abstr. 14th Int. Congr. Biochem., Prague, Czechoslovakia. **1988,** *5*, 47.
57. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991,** *37*, 487-493.
58. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991,** *354*, 82-84.
59. Geysen, H. M.; Rodda, S. J.; Mason, T. J. *Mol. Immunol.* **1986,** *23*, 709-715.
60. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991,** *354*, 84-86.
61. Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *Biotechniques* **1992,** *13*, 901-905.
62. Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*, 5381-5383.
63. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90*, 10700-10704.
64. Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993,** *115*, 9812-9813.
65. Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Neri, D. *J. Biotechnol.* **2006,** *126*, 568-581.
66. Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004,** *22*, 568-574.
67. Franch, T., et al. Patent number WO2007062664, 2007.
68. Morgan, B., et al. Patent number WO2005058479, 2005.
69. Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994,** *59*, 4723-4724.
70. Kottani, R.; Valiulin, R. A.; Kutateladze, A. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 13917-13921.
71. Egner, B. J.; Rana, S.; Smith, H.; Bouloc, N.; Frey, J. G.; Brocklesby, W. S.; Bradley, M. *Chem. Commun.* **1997**, 735-736.

72. Pastor, J. J.; Lingard, I.; Bhalay, G.; Bradley, M. *J. Comb. Chem.* **2003,** *5*, 85-90.
73. Ede, N. J.; Wu, Z. *Curr. Opin. Chem. Biol.* **2003,** *7*, 374-379.
74. Braeckmans, K.; De Smedt, S. C.; Leblans, M.; Pauwels, R.; Demeester, J. *Nat. Rev. Drug Discov.* **2002,** *1*, 447-456.
75. Battersby, B. J.; Bryant, D.; Meutermans, W.; Matthews, D.; Smythe, M. L.; Trau, M. *J. Am. Chem. Soc.* **2000,** *122*, 2138-2139.
76. Pianowski, Z. L.; Winssinger, N. *Chem. Soc. Rev.* **2008,** *37*, 1330-1336.
77. Harris, J. L.; Winssinger, N. *Chem. Eur. J.* **2005,** *11*, 6792-6801.
78. Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2001,** *40*, 3152-3155.
79. Winssinger, N.; Ficarro, S.; Schultz, P. G.; Harris, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 11139-11144.
80. Urbina, H. D.; Debaene, F.; Jost, B.; Bole-Feysot, C.; Mason, D. E.; Kuzmic, P.; Harris, J. L.; Winssinger, N. *ChemBioChem* **2006,** *7*, 1790-1797.
81. Diaz-Mochon, J. J.; Bialy, L.; Keinicke, L.; Bradley, M. *Chem. Commun.* **2005**, 1384-1386.
82. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M. *Chem. Commun.* **2006**, 3984-3986.
83. Stears, R. L.; Martinsky, T.; Schena, M. *Nature Med.* **2003,** *9*, 140-145.
84. Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995,** *270*, 467-470.
85. MacBeath, G.; Schreiber, S. L. *Science* **2000,** *289*, 1760-1763.
86. Houseman, B. T.; Huh, J. H.; Kron, S. J.; Mrksich, M. *Nat. Biotechnol.* **2002,** *20*, 270-274.
87. Kononen, J.; Bubendorf, L.; Kallioniemi, A.; Barlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P. *Nat. Med.* **1998,** *4*, 844-847.
88. Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A. *Nat. Biotechnol.* **2002,** *20*, 275-281.
89. Zlauddin, J.; Sabatini, D. M. *Nature* **2001,** *411*, 107-110.
90. MacBeath, G.; Koehler, A. N.; Schreiber, S. L. *J. Am. Chem. Soc.* **1999,** *121*, 7967-7968.
91. Mant, A.; Tourniaire, G.; Diaz-Mochon, J. J.; Elliott, T. J.; Williams, A. P.; Bradley, M. *Biomaterials* **2006,** *27*, 5299-5306.
92. Schena, M. *Microarray Analysis; New York: Wiley-Liss, John Wiley & Sons, Inc.* **2003**, 95-120.
93. Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91*, 5022-5026.
94. Pirrung, M. C. *Angew. Chem., Int. Ed.* **2002,** *41*, 1276-1289.
95. Hughes, T. R., et al. *Nat. Biotechnol.* **2001,** *19*, 342-347.
96. Brown, P. O.; Botstein, D. *Nat. Genet.* **1999,** *21*, 33-37.
97. Gresham, D.; Dunham, M. J.; Botstein, D. *Nat. Rev. Genet.* **2008,** *9*, 291-302.
98. Pollack, J. R.; Perou, C. M.; Alizadeh, A. A.; Eisen, M. B.; Pergamenschikov, A.; Williams, C. F.; Jeffrey, S. S.; Botstein, D.; Brown, P. O. *Nat. Genet.* **1999,** *23*, 41-46.
99. Henderson, G.; Bradley, M. *Curr. Opin. Biotech.* **2007,** *18*, 326-330.
100. Panicker, R. C.; Huang, X.; Yao, S. Q. *Comb. Chem. High Throughput Screen.* **2004,** *7*, 547-556.

101. Fodor, S. P.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991,** *251*, 767-773.
102. Schutkowski, M.; Reimer, U.; Panse, S.; Dong, L.; Lizcano, J. M.; Alessi, D. R.; Schneider-Mergener, J. *Angew. Chem., Int. Ed.* **2004,** *43*, 2671-2674.
103. Koehn, M.; Gutierrez-Rodriguez, M.; Jonkheijm, P.; Wetzel, S.; Wacker, R.; Schroeder, H.; Prinz, H.; Niemeyer, C. M.; Breinbauer, R.; Szedlacsek, S. E.; Waldmann, H. *Angew. Chem., Int. Ed.* **2007,** *46*, 7700-7703.
104. Sun, H.; Lu, C. H. S.; Uttamchandani, M.; Xia, Y.; Liou, Y.-C.; Yao, S. Q. *Angew. Chem., Int. Ed.* **2008,** *47*, 1698-1702.
105. Salisbury, C. M.; Maly, D. J.; Ellman, J. A. *J. Am. Chem. Soc.* **2002,** *124*, 14868-14870.
106. Reineke, U.; Ivascu, C.; Schlief, M.; Landgraf, C.; Gericke, S.; Zahn, G.; Herzel, H.; Volkmer-Engert, R.; Schneider-Mergener, J. *J. Immunol. Methods* **2002,** *267*, 37-51.
107. Andresen, H.; Groetzinger, C.; Zarse, K.; Kreuzer, O. J.; Ehrentreich-Foerster, E.; Bier, F. F. *Proteomics* **2006,** *6*, 1376-1384.
108. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002,** *298*, 1912-1934.
109. Besant, P. G.; Tan, E.; Attwood, P. V. *Int. J. Biochem. Cell Biol.* **2003,** *35*, 297-309.
110. Tonks, N. K. *Nat. Rev. Mol. Cell Biol.* **2006,** *7*, 833-846.
111. Hanks, S. K.; Quinn, A. M.; Hunter, T. *Science* **1988,** *241*, 42-52.
112. Blume-Jensen, P.; Hunter, T. *Nature* **2001,** *411*, 355-365.
113. Cohen, P. *Nat. Rev. Drug Discov.* **2002,** *1*, 309-315.
114. Goldstein, D. M.; Gray, N. S.; Zarrinkar, P. P. *Nat. Rev. Drug Discov.* **2008,** *7*, 391-397.
115. Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Science* **2004,** *303*, 1800-1805.
116. Asano, T.; Ikegaki, I.; Satoh, S.-I.; Seto, M.; Sasaki, Y. *Cardiovasc. Drug Rev.* **1998,** *16*, 76-87.
117. Sehgal, S. N. *Clin. Biochem.* **1998,** *31*, 335-340.
118. Druker, B. J.; Tamura, S.; Buchdunger, E.; Ohno, S.; Segal, G. M.; Fanning, S.; Zimmermann, J.; Lydon, N. B. *Nat. Med.* **1996,** *2*, 561-566.
119. Barker, A. J.; Gibson, K. H.; Grundy, W.; Godfrey, A. A.; Barlow, J. J.; Healy, M. P.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Scarlett, L.; Henthorn, L.; Richards, L. *Bioorg. Med. Chem. Lett.* **2001,** *11*, 1911-1914.
120. Baselga, J. *Science* **2006,** *312*, 1175-1178.
121. Apsel, B.; Blair, J. A.; Gonzalez, B.; Nazif, T. M.; Feldman, M. E.; Aizenstein, B.; Hoffman, R.; Williams, R. L.; Shokat, K. M.; Knight, Z. A. *Nat. Chem. Biol.* **2008,** *4*, 691-699.
122. Ma, H.; Deacon, S.; Horiuchi, K. *Expert Opin. Drug Discov.* **2008,** *3*, 607-621.
123. Ishida, A.; Kameshita, I.; Sueyoshi, N.; Taniguchi, T.; Shigeri, Y. *J. Pharmacol. Sci.* **2007,** *103*, 5-11.
124. Von Ahsen, O.; Bomer, U. *ChemBioChem* **2005,** *6*, 481-490.
125. Glickman, J. F.; Schmid, A.; Ferrand, S. *Assay Drug Dev. Technol.* **2008,** *6*, 433-455.

126. Lesaicherre, M.-L.; Uttamchandani, M.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2002,** *12*, 2085-2088.
127. Xu, K.; Stern, A. S.; Levin, W.; Chua, A.; Vassilev, L. T. *J. Biochem. Mol. Biol.* **2003,** *36*, 421-425.
128. Riddle, S. M.; Vedvik, K. L.; Hanson, G. T.; Vogel, K. W. *Anal. Biochem.* **2006,** *356*, 108-116.
129. Seethala, R. *Methods* **2000,** *22*, 61-70.
130. Koresawa, M.; Okabe, T. *Assay Drug Dev. Technol.* **2004,** *2*, 153-160.
131. Rininsland, F.; Xia, W.; Wittenburg, S.; Shi, X.; Stankewicz, C.; Achyuthan, K.; McBranch, D.; Whitten, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 15295-15300.
132. Songyang, Z.; Carraway, K. L.; Eck, M. J.; Harrison, S. C.; Feldman, R. A.; Mohammadi, M.; Schlessinger, J.; Hubbard, S. R.; Smith, D. P.; et al. *Nature* **1995,** *373*, 536-539.
133. Edman, P. *Nature* **1956,** *177*, 667-668.
134. Hutti, J. E.; Jarrell, E. T.; Chang, J. D.; Abbott, D. W.; Storz, P.; Toker, A.; Cantley, L. C.; Turk, B. E. *Nat. Methods* **2004,** *1*, 27-29.
135. Cujec, T. P.; Medeiros, P. F.; Hammond, P.; Rise, C.; Kreider, B. L. *Chem. Biol.* **2002,** *9*, 253-264.
136. Lam, K. S.; Lebl, M.; Krchnak, V. *Chem. Rev.* **1997,** *97*, 411-448.
137. Kim, Y.-G.; Shin, D.-S.; Kim, E.-M.; Park, H.-Y.; Lee, C.-S.; Kim, J.-H.; Lee, B.-S.; Lee, Y.-S.; Kim, B.-G. *Angew. Chem., Int. Ed.* **2007,** *46*, 5408-5411.
138. Shin, D.-S.; Kim, Y.-G.; Kim, E.-M.; Kim, M.; Park, H.-Y.; Kim, J.-H.; Lee, B.-S.; Kim, B.-G.; Lee, Y.-S. *J. Comb. Chem.* **2008,** *10*, 20-23.
139. Wu, J.; Ma, Q. N.; Lam, K. S. *Biochemistry* **1994,** *33*, 14825-14833.
140. Shults, M. D.; Kozlov, I. A.; Nelson, N.; Kermani, B. G.; Melnyk, P. C.; Shevchenko, V.; Srinivasan, A.; Musmacker, J.; Hachmann, J. P.; Barker, D. L.; Lebl, M.; Zhao, C. *ChemBioChem* **2007,** *8*, 933-942.
141. Schutkowski, M.; Reineke, U.; Reimer, U. *ChemBioChem* **2005,** *6*, 513-521.
142. Min, D.-H.; Su, J.; Mrksich, M. *Angew. Chem., Int. Ed.* **2004,** *43*, 5973-5977.
143. Su, J.; Bringer, M. R.; Ismagilov, R. F.; Mrksich, M. *J. Am. Chem. Soc.* **2005,** *127*, 7280-7281.
144. Kim, Y.-P.; Oh, E.; Oh, Y.-H.; Moon, D. W.; Lee, T. G.; Kim, H.-S. *Angew. Chem., Int. Ed.* **2007,** *46*, 6816-6819.
145. Panse, S.; Dong, L.; Burian, A.; Carus, R.; Schutkowski, M.; Reimer, U.; Schneider-Mergener, J. *Mol. Divers.* **2004,** *8*, 291-299.
146. Rychlewski, L.; Kschischo, M.; Dong, L.; Schutkowski, M.; Reimer, U. *J. Mol. Biol.* **2004,** *336*, 307-311.
147. Maller, J. L.; Kemp, B. E.; Krebs, E. G. *Proc. Natl. Acad. Sci. U.S.A.* **1978,** *75*, 248-251.
148. Lacombe, J.-M.; Andriamanampisoa, F.; Pavia, A. A. *Int. J. Pept. Protein Res.* **1990,** *36*, 275-280.
149. Wakamiya, T.; Saruta, K.; Yasuoka, J.; Kusumoto, S. *Chem. Lett.* **1994**, 1099-1102.
150. Perich, J. W.; Ede, N. J.; Eagle, S.; Bray, A. M. *Lett. Pept. Sci.* **1999,** *6*, 91-97.

151. Fischer, R.; Mader, O.; Jung, G.; Brock, R. *Bioconjugate Chem.* **2003,** *14*, 653-660.
152. Mujumdar, R. B.; Ernst, L. A.; Mujumdar, S. R.; Lewis, C. J.; Waggoner, A. S. *Bioconjugate Chem.* **1993,** *4*, 105-111.
153. Horiuchi, K. Y.; Wang, Y.; Diamond, S. L.; Ma, H. *J. Biomol. Screen.* **2006,** *11*, 48-56.
154. Martin, K.; Steinberg, T. H.; Cooley, L. A.; Gee, K. R.; Beechem, J. M.; Patton, W. F. *Proteomics* **2003,** *3*, 1244-1255.
155. Akita, S.; Umezawa, N.; Higuchi, T. *Org. Lett.* **2005,** *7*, 5565-5568.
156. Poot, A. J.; Ruijter, E.; Nuijens, T.; Dirksen, E. H. C.; Heck, A. J. R.; Slijper, M.; Rijkers, D. T. S.; Liskamp, R. M. J. *Proteomics* **2006,** *6*, 6394-6399.
157. Winssinger, N.; Damoiseaux, R.; Tully, D. C.; Geierstanger, B. H.; Burdick, K.; Harris, J. L. *Chem. Biol.* **2004,** *11*, 1351-1360.
158. Pouchain, D.; Diaz-Mochon, J. J.; Bialy, L.; Bradley, M. *ACS Chem. Biol.* **2007,** *2*, 810-818.
159. Weiler, J.; Gausepohl, H.; Hauser, N.; Jensen, O. N.; Hoheisel, J. D. *Nucleic Acids Res.* **1997,** *25*, 2792-2799.
160. Pendergast, A. M. *Adv. Cancer Res.* **2002,** *85*, 51-100.
161. Beier, M.; Hoheisel, J. D. *Nucleic Acids Res.* **1999,** *27*, 1970-1977.
162. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997,** *25*, 3389-3402.
163. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.-C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. *Nucleic Acids Res.* **2003,** *31*, 365-370.
164. Karlin, S.; Altschul, S. F. *Proc. Natl. Acad. Sci. U.S.A.* **1990,** *87*, 2264-2268.
165. Till, J. H.; Chan, P. M.; Miller, W. T. *J. Biol. Chem.* **1999,** *274*, 4995-5003.
166. Rowley, J. D. *Cancer* **1990,** *65*, 2178-2184.
167. Quintas-Cardama, A.; Kantarjian, H.; Cortes, J. *Nat. Rev. Drug Discov.* **2007,** *6*, 834-848.
168. Wu, J. J.; Afar, D. E. H.; Phan, H.; Witte, O. N.; Lam, K. S. *Comb. Chem. High Throughput Screen.* **2002,** *5*, 83-91.
169. Yoshida, K.; Yamaguchi, T.; Natsume, T.; Kufe, D.; Miki, Y. *Nat. Cell Biol.* **2005,** *7*, 278-285.
170. Tanos, B.; Pendergast, A. M. *J. Biol. Chem.* **2006,** *281*, 32714-32723.
171. Durig, J.; Schmucker, U.; Duhrsen, U. *Leukemia* **2001,** *15*, 752-756.
172. Ushio-Fukai, M.; Zuo, L.; Ikeda, S.; Tojo, T.; Patrushev, N. A.; Alexander, R. W. *Circ. Res.* **2005,** *97*, 829-836.
173. Yamada, K.; Sakane, F.; Imai Shin, i.; Tsushima, S.; Murakami, T.; Kanoh, H. *Biochem. Biophys. Res. Commun.* **2003,** *305*, 101-107.
174. Dorsam, R. T.; Gutkind, J. S. *Nat. Rev. Cancer* **2007,** *7*, 79-94.
175. Moasser, M. M. *Oncogene* **2007,** *26*, 6469-6487.
176. Hayes, D. F., et al. *N. Engl. J. Med.* **2007,** *357*, 1496-1506.
177. Carter, P.; Presta, L.; Gorman, C. M.; Ridgway, J. B.; Henner, D.; Wong, W. L.; Rowland, A. M.; Kotts, C.; Carver, M. E.; Shepard, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*, 4285-4289.
178. Geyer, C. E., et al. *N. Engl. J. Med.* **2006,** *355*, 2733-2743.
179. Eller, M.; Jarv, J.; Toomik, R.; Ragnarsson, U.; Ekman, P.; Engstroem, L. *J. Biochem.* **1993,** *114*, 177-180.

180. Chan, P. M.; Nestler, H. P.; Miller, W. T. *Cancer Lett.* **2000,** *160*, 159-169.
181. Sen, S.; Zhou, H.; White, R. A. *Oncogene* **1997,** *14*, 2195-2200.
182. Fox, B. P.; Kandpal, R. P. *Biochem. Biophys. Res. Commun.* **2004,** *318*, 882-892.
183. Negro, A.; Brar, B. K.; Gu, Y.; Peterson, K. L.; Vale, W.; Lee, K.-F. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 15889-15893.
184. Olsson, A.-K.; Dimberg, A.; Kreuger, J.; Claesson-Welsh, L. *Nat. Rev. Mol. Cell Biol.* **2006,** *7*, 359-371.
185. Ferrara, N.; Hillan, K. J.; Gerber, H.-P.; Novotny, W. *Nat. Rev. Drug Discov.* **2004,** *3*, 391-400.
186. Sun, L., et al. *J. Med. Chem.* **2003,** *46*, 1116-1119.
187. Atkins, M.; Jones, C. A.; Kirkpatrick, P. *Nat. Rev. Drug Discov.* **2006,** *5*, 279-280.
188. Wilhelm, S.; Carter, C.; Lynch, M. P.; Lowinger, T.; Dumas, J.; Smith, R. A.; Schwartz, B.; Simantov, R.; Kelley, S. *Nat. Rev. Drug Discov.* **2006,** *5*, 835-844.
189. Gonzalez, R. R.; Cherfils, S.; Escobar, M.; Yoo, J. H.; Carino, C.; Styer, A. K.; Sullivan, B. T.; Sakamoto, H.; Olawaiye, A.; Serikawa, T.; Lynch, M. P.; Rueda, B. R. *J. Biol. Chem.* **2006,** *281*, 26320-26328.
190. Singh, A. J.; Meyer, R. D.; Band, H.; Rahimi, N. *Mol. Biol. Cell* **2005,** *16*, 2106-2118.
191. Endo, A.; Nagashima, K.-I.; Kurose, H.; Mochizuki, S.; Matsuda, M.; Mochizuki, N. *J. Biol. Chem.* **2002,** *277*, 23747-23754.
192. Puente, X. S.; Sanchez, L. M.; Overall, C. M.; Lopez-Otin, C. *Nat. Rev. Genet.* **2003,** *4*, 544-558.
193. Barrett, A. J.; Rawlings, N. D.; Woessner, J. F. *Handbook of Proteolytic Enzymes, 2nd ed.; Academic Press: 2004*.
194. Leung, D.; Abbenante, G.; Fairlie, D. P. *J. Med. Chem.* **2000,** *43*, 305-341.
195. Turk, B. *Nat. Rev. Drug Discovery* **2006,** *5*, 785-799.
196. Nazif, T.; Bogyo, M. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 2967-2972.
197. Petithory, J. R.; Masiarz, F. R.; Kirsch, J. F.; Santi, D. V.; Malcolm, B. A. *Proc. Natl. Acad. Sci. U. S. A.* **1991,** *88*, 11510-11514.
198. Turk, B. E.; Huang, L. L.; Piro, E. T.; Cantley, L. C. *Nat. Biotechnol.* **2001,** *19*, 661-667.
199. Backes, B. J.; Harris, J. L.; Leonetti, F.; Craik, C. S.; Ellman, J. A. *Nat. Biotechnol.* **2000,** *18*, 187-193.
200. Thornberry, N. A.; Rano, T. A.; Peterson, E. P.; Rasper, D. M.; Timkey, T.; Garcia-Calvo, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; Chapman, K. T.; Nicholson, D. W. *J. Biol. Chem.* **1997,** *272*, 17907-17911.
201. Harris, J. L.; Backes, B. J.; Leonetti, F.; Mahrus, S.; Ellman, J. A.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 7754-7759.
202. Debaene, F.; Mejias, L.; Harris, J. L.; Winssinger, N. *Tetrahedron* **2004,** *60*, 8677-8690.
203. Harris, J.; Mason, D. E.; Li, J.; Burdick, K. W.; Backes, B. J.; Chen, T.; Shipway, A.; Van Heeke, G.; Gough, L.; Ghaemmaghami, A.; Shakib, F.; Debaene, F.; Winssinger, N. *Chem. Biol.* **2004,** *11*, 1361-1372.

204. Choe, Y.; Leonetti, F.; Greenbaum, D. C.; Lecaille, F.; Bogyo, M.; Broemme, D.; Ellman, J. A.; Craik, C. S. *J. Biol. Chem.* **2006,** *281*, 12824-12832.
205. Barrios, A. M.; Craik, C. S. *Bioorg. Med. Chem. Lett.* **2002,** *12*, 3619-3623.
206. Gosalia, D. N.; Salisbury, C. M.; Maly, D. J.; Ellman, J. A.; Diamond, S. L. *Proteomics* **2005,** *5*, 1292-1298.
207. Gosalia, D. N.; Salisbury, C. M.; Ellman, J. A.; Diamond, S. L. *Mol. Cell. Proteomics* **2005,** *4*, 626-636.
208. Angenendt, P.; Lehrach, H.; Kreutzberger, J.; Gloekler, J. *Proteomics* **2005,** *5*, 420-425.
209. Knight, C. G. *Methods Enzymol.* **1995,** *248*, 18-34.
210. Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1994,** *91*, 3314-3318.
211. Rosse, G.; Kueng, E.; Page, M. G.; Schauer-Vukasinovic, V.; Giller, T.; Lahm, H. W.; Hunziker, P.; Schlatter, D. *J. Comb. Chem.* **2000,** *2*, 461-466.
212. Petrassi, H. M.; Williams, J. A.; Li, J.; Tumanut, C.; Ek, J.; Nakai, T.; Masick, B.; Backes, B. J.; Harris, J. L. *Bioorg. Med. Chem. Lett.* **2005,** *15*, 3162-3166.
213. Sano, S.; Tomizaki, K.-Y.; Usui, K.; Mihara, H. *Bioorg. Med. Chem. Lett.* **2006,** *16*, 503-506.
214. Meldal, M. *Tetrahedron Lett.* **1992,** *33*, 3077-3080.
215. Rink, H. *Tetrahedron Lett.* **1987,** *28*, 3787-3790.
216. Chhabra, S. R.; Parekh, H.; Khan, A. N.; Bycroft, B. W.; Kellam, B. *Tetrahedron Lett.* **2001,** *42*, 2189-2192.
217. http://www.appliedbiosystems.com/support/seqguide.cfm, accessed the 11th February 2009.
218. Chhabra, S. R.; Hothi, B.; Evans, D. J.; White, P. D.; Bycroft, B. W.; Chan, W. C. *Tetrahedron Lett.* **1998,** *39*, 1603-1606.
219. Zhao, Z. G.; Im, J. S.; Lam, K. S.; Lake, D. F. *Bioconjugate Chem.* **1999,** *10*, (3), 424-430.
220. Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970,** *34*, 595-598.
221. Ternon, M.; Diaz-Mochon, J. J.; Belsom, A.; Bradley, M. *Tetrahedron* **2004,** *60*, 8721-8728.
222. Thieriet, N.; Alsina, J.; Giralt, E.; Guibe, F.; Albericio, F. *Tetrahedron Lett.* **1997,** *38*, 7275-7278.
223. Brenner, S.; Williams, S. R.; Vermaas, E. H.; Storck, T.; Moon, K.; McCollum, C.; Mao, J. I.; Luo, S.; Kirchner, J. J.; Eletr, S.; DuBridge, R. B.; Burcham, T.; Albrecht, G. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97*, 1665-1670.
224. Altenbrunn, F.; Seitz, O. *Org. Biomol. Chem.* **2008,** *6*, 2493-2498.
225. Fruton, J. S. *Q. Rev. Biol.* **2002,** *77*, 127-147.
226. Linderstrom-Lang, K.; Ottesen, M. *Nature* **1947,** *159*, 807-808.
227. Grahn, S.; Ullmann, D.; Jakubke, H. D. *Anal. Biochem.* **1998,** *265*, 225-231.
228. Diaz-Mochon, J. J.; Planonth, S.; Bradley, M. *Anal. Biochem.* **2009,** *384*, 101-105.
229. Fields, G. B.; Noble, R. L. *Int. J. Pept. Protein Res.* **1990,** *35*, 161-214.
230. Sarin, V. K.; Kent, S. B.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981,** *117*, 147-157.
231. Vojkovsky, T. *Pept. Res.* **1995,** *8*, 236-237.

APPENDICES

Appendix I: Sequences of the 155 "hits" identified from the Abl screening. The one-letter amino-acid code is used.

1) FQSEYEVIK	33) FQSEYIPIK	65) FQKEYSEIK	97) FQPKYEEIK	129) FQvEYRVIK
2) FQSEYESIK	34) FQEEYESIK	66) FQVAYSSIK	98) FQEEYSVIK	130) FQPSYSEIK
3) FQSEYSFIK	35) FQSEYKPIK	67) FQSEYKIIK	99) FQvEYKSIK	131) FQRVYSKIK
4) FQSEYVFIK	36) FQpEYSSIK	68) FQvEYEVIK	100) FQKVYERIK	132) FQvSYSVIK
5) FQVIYESIK	37) FQpEYVEIK	69) FQPEYKFIK	101) FQESYEVIK	133) FQESYSPIK
6) FQSIYEPIK	38) FQPRYSSIK	70) FQPSYESIK	102) FQREYKIIK	134) FQSSYEVIK
7) FQSIYSPIK	39) FQSEYVIIK	71) FQSEYVRIK	103) FQSEYVvIK	135) FQEKYSVIK
8) FQPEYSEIK	40) FQKVYSEIK	72) FQKVYEEIK	104) FQFSYFIIK	136) FQSKYSVIK
9) FQSEYEEIK	41) FQSVYEEIK	73) FQEEYSRIK	105) FQvEYFKIK	137) FQEPYEEIK
10) FQVEYESIK	42) FQVEYSVIK	74) FQSVYEPIK	106) FQvEYpKIK	138) FQEKYSEIK
11) FQSEYSSIK	43) FQvEYFFIK	75) FQpVYSSIK	107) FQEEYKKIK	139) FQvEYKpIK
12) FQKIYSVIK	44) FQVEYSSIK	76) FQKVYEFIK	108) FQSVYKVIK	140) FQESYSKIK
13) FQPEYESIK	45) FQPAYESIK	77) FQKEYFVIK	109) FQvEYRSIK	141) FQvEYRKIK
14) FQEVYSSIK	46) FQEVYKVIK	78) FQvEYKVIK	110) FQvVYSKIK	142) FQEEYKVIK
15) FQRIYSEIK	47) FQSEYERIK	79) FQSEYVKIK	111) FQPKYEVIK	143) FQKEYSKIK
16) FQPEYVVIK	48) FQpEYVVIK	80) FQvVYSPIK	112) FQpVYEKIK	144) FQEEYKpIK
17) FQSEYVSIK	49) FQEEYEVIK	81) FQvEYVSIK	113) FQvEYRpIK	145) FQSKYEEIK
18) FQEEYVFIK	50) FQpEYKVIK	82) FQKEYEEIK	114) FQSEYRVIK	146) FQEaYSSIK
19) FQEVYSFIK	51) FQEVYEEIK	83) FQESYSVIK	115) FQvEYSRIK	147) FQEKYEPIK
20) FQSEYEFIK	52) FQSVYSPIK	84) FQVEYKPIK	116) FQSEYEpIK	148) FQKEYEKIK
21) FQPIYRSIK	53) FQpEYISIK	85) FQSRYESIK	117) FQREYSKIK	149) FQpEYKKIK
22) FQEEYSPIK	54) FQRVYSVIK	86) FQRVYSSIK	118) FQvEYEvIK	150) FQEEYSvIK
23) FQKIYSEIK	55) FQEVYFVIK	87) FQEEYSpIK	119) FQVSYSSIK	151) FQVSYEKIK
24) FQRVYSEIK	56) FQREYFSIK	88) FQKVYESIK	120) FQSSYEEIK	152) FQvEYvRIK
25) FQPVYEKIK	57) FQvEYEFIK	89) FQSSYSEIK	121) FQpEYEpIK	153) FQEaYESIK
26) FQEEYSFIK	58) FQKEYFPIK	90) FQvVYERIK	122) FQKEYESIK	154) FQvSYEvIK
27) FQSEYFFIK	59) FQREYSFIK	91) FQPAYSSIK	123) FQvEYVvIK	155) FQKPYEVIK
28) FQpEYESIK	60) FQvEYIEIK	92) FQPEYKRIK	124) FQESYESIK	
29) FQvEYSPIK	61) FQpVYSVIK	93) FQpEYKSIK	125) FQPEYKKIK	
30) FQKIYEKIK	62) FQSAYSSIK	94) FQvEYSvIK	126) FQvEYKRIK	
31) FQSEYEIIK	63) FQFEYSSIK	95) FQEVYKIIK	127) FQpSYSVIK	
32) FQSEYEKIK	64) FQSVYSFIK	96) FQSVYSKIK	128) FQSSYVPIK	

Appendix II (first part): Sequences of the 299 "hits" identified from the Her2 screening. The one-letter amino-acid code is used.

1) FQpRYPpIK	39) FQKIYSRIK	77) FQSEYpSIK	115) FQVPYEFIK	153) FQKaYSSIK
2) FQVRYFpIK	40) FQVKYSPIK	78) FQpAYpSIK	116) FQKSYpSIK	154) FQKaYFpIK
3) FQFaYppIK	41) FQPRYvSIK	79) FQSaYFpIK	117) FQRaYpSIK	155) FQPSYEpIK
4) FQpIYppIK	42) FQRSYSEIK	80) FQSRYpFIK	118) FQEKYSpIK	156) FQvRYvSIK
5) FQpAYpRIK	43) FQEAYppIK	81) FQSEYFpIK	119) FQvEYvKIK	157) FQKSYRpIK
6) FQKIYppIK	44) FQKaYPpIK	82) FQKVYIPIK	120) FQVaYRPIK	158) FQREYKPIK
7) FQFAYpSIK	45) FQFRYKSIK	83) FQvSYRpIK	121) FQKRYEFIK	159) FQESYEpIK
8) FQPAYFpIK	46) FQRaYSpIK	84) FQSRYSEIK	122) FQRKYvpIK	160) FQEVYFSIK
9) FQEAYRSIK	47) FQREYpvIK	85) FQvSYRIK	123) FQpVYVKIK	161) FQVIYvKIK
10) FQIaYppIK	48) FQRSYRRIK	86) FQRSYvSIK	124) FQEVYEKIK	162) FQpaYVKIK
11) FQPRYSpIK	49) FQKEYIpIK	87) FQVSYSRIK	125) FQKSYFSIK	163) FQIaYSSIK
12) FQISYppIK	50) FQpKYSSIK	88) FQSEYIEIK	126) FQpSYppIK	164) FQVEYvpIK
13) FQRSYFpIK	51) FQKRYppIK	89) FQvRYPpIK	127) FQFKYSpIK	165) FQvSYPpIK
14) FQVRYRSIK	52) FQKAYSpIK	90) FQRKYppIK	128) FQSPYSpIK	166) FQRaYPpIK
15) FQSIYRSIK	53) FQISYSpIK	91) FQVaYppIK	129) FQEKYKpIK	167) FQpKYPVIK
16) FQEIYKRIK	54) FQERYvpIK	92) FQSaYpSIK	130) FQKRYESIK	168) FQpaYSpIK
17) FQpRYREIK	55) FQKRYFSIK	93) FQRKYpvIK	131) FQSVYFpIK	169) FQESYSRIK
18) FQpRYPSIK	56) FQSAYSPIK	94) FQERYvSIK	132) FQSaYSpIK	170) FQpSYvSIK
19) FQSRYppIK	57) FQvSYppIK	95) FQKSYpKIK	133) FQKSYEKIK	171) FQKSYFpIK
20) FQFSYERIK	58) FQRRYvRIK	96) FQSSYSSIK	134) FQFEYEKIK	172) FQvaYEpIK
21) FQvaYFpIK	59) FQSSYvpIK	97) FQVKYSpIK	135) FQpAYSSIK	173) FQSKYpSIK
22) FQFaYSpIK	60) FQpSYPpIK	98) FQERYFSIK	136) FQSEYvpIK	174) FQvSYSpIK
23) FQRRYpSIK	61) FQISYpSIK	99) FQVSYKpIK	137) FQKSYSpIK	175) FQPSYEPIK
24) FQIaYPpIK	62) FQREYEpIK	100) FQREYppIK	138) FQpSYSpIK	176) FQvEYpSIK
25) FQSRYPSIK	63) FQVSYREIK	101) FQpAYEKIK	139) FQEEYppIK	177) FQFSYvSIK
26) FQEaYppIK	64) FQSRYpSIK	102) FQPKYvpIK	140) FQVEYSFIK	178) FQSSYRSIK
27) FQIVYSSIK	65) FQpEYPpIK	103) FQFKYpSIK	141) FQSAYvSIK	179) FQKSYSEIK
28) FQRAYpPIK	66) FQKVYFpIK	104) FQERYpSIK	142) FQVaYpSIK	180) FQpEYpSIK
29) FQPSYFSIK	67) FQpPYppIK	105) FQvSYFpIK	143) FQSaYvpIK	181) FQEEYPpIK
30) FQvaYFSIK	68) FQPIYIVIK	106) FQvRYpSIK	144) FQPSYpSIK	182) FQKVYppIK
31) FQIEYSRIK	69) FQERYFFIK	107) FQVPYRpIK	145) FQKPYpFIK	183) FQVVYSSIK
32) FQISYKRIK	70) FQFEYpSIK	108) FQISYvSIK	146) FQKRYSpIK	184) FQESYvEIK
33) FQISYRpIK	71) FQvRYVFIK	109) FQKSYSFIK	147) FQEKYvpIK	185) FQpAYvSIK
34) FQvAYvpIK	72) FQVSYppIK	110) FQvRYppIK	148) FQRKYPSIK	186) FQpRYvKIK
35) FQIPYvpIK	73) FQKRYpSIK	111) FQEVYVVIK	149) FQpSYFSIK	187) FQFSYSpIK
36) FQRaYVRIK	74) FQSRYSpIK	112) FQSRYvSIK	150) FQEaYpFIK	188) FQSaYpvIK
37) FQSIYPpIK	75) FQESYIRIK	113) FQPSYvpIK	151) FQKaYKSIK	189) FQSEYPpIK
38) FQIaYpSIK	76) FQVEYKVIK	114) FQVSYRpIK	152) FQPaYSEIK	190) FQESYPpIK

Appendix II (second part): Sequences of the 299 “hits” identified from the Her2 screening. The one-letter amino-acid code is used.

191) FQVKYRSIK	213) FQVSYEEIK	235) FQvSYvSIK	257) FQKKYppIK	279) FQSKYSvIK
192) FQpaYRKIK	214) FQEKYvSIK	236) FQSSYPSIK	258) FQRKYEKIK	280) FQpKYpSIK
193) FQEPYSSIK	215) FQvaYvSIK	237) FQSSYKSIK	259) FQKRYpKIK	281) FQKEYEPIK
194) FQvEYvSIK	216) FQEEYRpIK	238) FQVVYPKIK	260) FQPSYvSIK	282) FQEEYvSIK
195) FQSaYSPIK	217) FQESYRFIK	239) FQREYvSIK	261) FQEKYppIK	283) FQFKYIvIK
196) FQSKYvpIK	218) FQKKYvFIK	240) FQVaYKSIK	262) FQERYRvIK	284) FQpEYSSIK
197) FQFaYvSIK	219) FQKSYvSIK	241) FQpSYSSIK	263) FQSVYvSIK	285) FQEKYSEIK
198) FQEKYFEIK	220) FQESYSpIK	242) FQVKYpvIK	264) FQKEYpSIK	286) FQKSYpvIK
199) FQVVYpKIK	221) FQpEYvpIK	243) FQvRYKVIK	265) FQKEYKpIK	287) FQEKYVRIK
200) FQvaYSSIK	222) FQRKYvKIK	244) FQKKYSKIK	266) FQKEYSFIK	288) FQvKYvRIK
201) FQKRYpEIK	223) FQEKYRpIK	245) FQRKYSpIK	267) FQKSYESIK	289) FQKKYvSIK
202) FQSSYSFIK	224) FQpaYSPIK	246) FQEaYKPIK	268) FQRKYpKIK	290) FQKSYpEIK
203) FQvAYKSIK	225) FQKaYKpIK	247) FQSaYKFIK	269) FQpPYKPIK	291) FQREYKSIK
204) FQpKYvSIK	226) FQKEYRSIK	248) FQKVYvpIK	270) FQvPYKpIK	292) FQKSYSvIK
205) FQRKYSSIK	227) FQKKYSvIK	249) FQvKYppIK	271) FQRaYpKIK	293) FQKKYpSIK
206) FQpEYKvIK	228) FQKSYPpIK	250) FQKEYSpIK	272) FQKVYEPIK	294) FQKaYvVIK
207) FQSEYppIK	229) FQSaYRPIK	251) FQKaYvRIK	273) FQEaYSEIK	295) FQvKYVPIK
208) FQPaYSSIK	230) FQSKYRFIK	252) FQvKYvpIK	274) FQKKYSFIK	296) FQvKYKpIK
209) FQPSYSSIK	231) FQSSYpRIK	253) FQIRYvVIK	275) FQEPYvEIK	297) FQEEYEvIK
210) FQpKYvKIK	232) FQSaYERIK	254) FQvEYpvIK	276) FQSaYKSIK	298) FQvKYpKIK
211) FQvEYSpIK	233) FQRaYSvIK	255) FQvKYRPIK	277) FQSKYvFIK	299) FQVKYKKIK
212) FQEVYEpIK	234) FQVKYSVIK	256) FQERYvKIK	278) FQvVYSEIK	

Appendix III: Sequences of the top 50 "hits" identified from the VEGFR2 screening. The one-letter amino-acid code is used.

1) FQIAYPFIK	14) FQPRYRSIK	27) FQpAYISIK	40) FQPRYpKIK
2) FQFRYEIIK	15) FQIVYVRIK	28) FQEAYEFIK	41) FQPRYpEIK
3) FQIAYPPIK	16) FQVAYIpIK	29) FQFEYPRIK	42) FQPEYESIK
4) FQFEYISIK	17) FQPEYIvIK	30) FQFRYPSIK	43) FQFIYRPIK
5) FQPAYIvIK	18) FQISYFSIK	31) FQIAYRSIK	44) FQVAYSSIK
6) FQIRYPPIK	19) FQIAYFEIK	32) FQIRYpEIK	45) FQIRYKIIK
7) FQEAYPEIK	20) FQFAYIpIK	33) FQpIYVFIK	46) FQRRYpSIK
8) FQIAYvpIK	21) FQFVYFRIK	34) FQRIYpRIK	47) FQIaYpFIK
9) FQFRYKFIK	22) FQPSYIpIK	35) FQpAYEEIK	48) FQPIYRpIK
10) FQIVYppIK	23) FQIIYSFIK	36) FQvIYPSIK	49) FQFRYEKIK
11) FQEIYIPIK	24) FQFIYEEIK	37) FQFKYFRIK	50) FQSIYFRIK
12) FQIVYREIK	25) FQFRYSvIK	38) FQFaYFPIK	
13) FQFSYFSIK	26) FQPIYpFIK	39) FQSAYSRIK	

Appendix IV: Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with pepsin (625-member PNA-encoded positional scanning library)

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Ala	0,134822872	0,002078703
Ser	0,102141838	0,001138134
Lys	0,113698102	0,003685247
Glu	0,064925363	0,000958684
Pro	0,083199791	0,001259414
Ala (duplicate)	0,105484549	0,002943523
Ser (duplicate)	0,077187209	0,001509803
Lys (duplicate)	0,088175009	0,001266744
Glu (duplicate)	0,054808912	0,000703658
Pro (duplicate)	0,079824635	0,001442406

Appendix V: Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with subtilisin A (625-member PNA-encoded positional scanning library)

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Ala	0,793850713	0,027925311
Ser	0,321207712	0,003470474
Lys	0,290136845	0,007538769
Glu	0,151069519	0,027133712
Pro	0,285344764	0,012224464
Ala (duplicate)	0,654410652	0,056794691
Ser (duplicate)	0,284242084	0,005633034
Lys (duplicate)	0,238276756	0,006686731
Glu (duplicate)	0,142058441	0,003398875
Pro (duplicate)	0,189564069	0,014303455

Appendix VI: Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with trypsin (625-member PNA-encoded positional scanning library)

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Ala	2,304500703	0,134153434
Ser	0,409974587	0,012058739
Lys	1,130846774	0,031892212
Glu	0,620923762	0,018718482
Pro	0,512729264	0,031587681
Ala (duplicate)	3,219626168	0,284608895
Ser (duplicate)	0,378708148	0,021960912
Lys (duplicate)	2,300243605	0,111898463
Glu (duplicate)	0,469775281	0,030903061
Pro (duplicate)	1,348504551	0,049651044

Appendix VII (first part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with trypsin (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Pro (1)	0,512684156	0,0053
Thr (1)	0,48671885	0,0086
Tyr (1)	0,414548052	0,0150
Arg (1)	0,504991979	0,0111
Glu (1)	0,499160291	0,0091
His (1)	0,440835229	0,0138
Lys (1)	0,424113348	0,0174
Phe (1)	0,4553458	0,0150
Ser (1)	0,432391637	0,0163
Trp (1)	0,455771576	0,0125
Val (1)	0,380207691	0,0143
Ala (1)	0,460480365	0,0104
Gln (1)	0,451866238	0,0353
Gly (1)	0,412047254	0,0100
Ile (1)	0,390723943	0,0117
Pro (2)	0,374823313	0,0098
Thr (2)	0,338079981	0,0075
Tyr (2)	0,314632632	0,0072
Arg (2)	0,456797099	0,0072
Glu (2)	0,352706506	0,0092
His (2)	0,326281815	0,0051
Lys (2)	0,311408269	0,0025
Phe (2)	0,427825754	0,0098
Ser (2)	0,33989809	0,0047

Appendix VII (second part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with trypsin (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Trp (2)	0,339506545	0,0058
Val (2)	0,305565278	0,0087
Ala (2)	0,448581288	0,0172
Gln (2)	0,401343468	0,0089
Gly (2)	0,337501961	0,0069
Ile (2)	0,315027097	0,0075
Pro (3)	0,589709763	0,0180
Thr (3)	0,515820943	0,0122
Tyr (3)	0,44923525	0,0135
Arg (3)	0,523689551	0,0095
Glu (3)	0,456875468	0,0108
His (3)	0,427128008	0,0151
Lys (3)	0,412917485	0,0105
Phe (3)	0,48071841	0,0134
Ser (3)	0,466614214	0,0185
Trp (3)	0,437041651	0,0162
Val (3)	0,404002931	0,0246
Ala (3)	0,522692721	0,0100
Gln (3)	0,484006314	0,0161
Gly (3)	0,463140179	0,0265
Ile (3)	0,408800612	0,0122
Pro (4)	0,542402361	0,0353
Thr (4)	0,455311364	0,0119
Tyr (4)	0,441425021	0,0143
Arg (4)	0,602178381	0,0193
Glu (4)	0,449454927	0,0175
His (4)	0,474646857	0,0190
Lys (4)	0,452389554	0,0204
Phe (4)	0,56370095	0,0219
Ser (4)	0,561056106	0,1051
Trp (4)	0,445637428	0,0166
Val (4)	0,452837893	0,0576
Ala (4)	0,515199884	0,0120
Gln (4)	0,46578856	0,0184
Gly (4)	0,478595924	0,0192
Ile (4)	0,45967812	0,0200

Appendix VIII (first part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with chymopapain (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Pro (1)	0,626989724	0,019162873
Thr (1)	0,667693523	0,052493697
Tyr (1)	0,560068216	0,024841138
Arg (1)	0,648314797	0,031734147
Glu (1)	0,750158529	0,121093727
His (1)	0,60725846	0,02075028
Lys (1)	0,601751556	0,01724031
Phe (1)	0,634275114	0,019159656
Ser (1)	0,609491236	0,011890696
Trp (1)	0,628397667	0,016061516
Val (1)	0,540541968	0,015891027
Ala (1)	0,651470225	0,01754194
Gln (1)	0,666253038	0,011225183
Gly (1)	0,623170116	0,006646447
Ile (1)	0,535624274	0,00610106
Pro (2)	0,459059944	0,036950833
Thr (2)	0,546473913	0,03675631
Tyr (2)	0,482407968	0,010073198
Arg (2)	0,502550277	0,026676672
Glu (2)	0,500126935	0,022452691
His (2)	0,488197882	0,008722781
Lys (2)	0,476258993	0,004674795
Phe (2)	0,500128524	0,011453786
Ser (2)	0,529582515	0,016959364
Trp (2)	0,499358999	0,011097825
Val (2)	0,465889481	0,008106327
Ala (2)	0,567223237	0,01180252
Gln (2)	0,540125535	0,035253149
Gly (2)	0,497513114	0,012878881
Ile (2)	0,458549343	0,012599655
Pro (3)	0,809107833	0,058996634
Thr (3)	0,699202668	0,034638241
Tyr (3)	0,607446552	0,003981597
Arg (3)	0,577773352	0,024795897
Glu (3)	0,65826002	0,020413562
His (3)	0,621856734	0,007350107
Lys (3)	0,613649188	0,019381016
Phe (3)	0,57931598	0,035526272
Ser (3)	0,684141693	0,019819054
Trp (3)	0,633067024	0,004793043
Val (3)	0,58375879	0,027415066
Ala (3)	0,627755616	0,037525647
Gln (3)	0,668483803	0,024038264
Gly (3)	0,67466708	0,010754464
Ile (3)	0,654049557	0,015451285
Pro (4)	0,714795244	0,089997482
Thr (4)	0,609141565	0,016836553
Tyr (4)	0,626196426	0,008233358

Appendix VIII (second part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with chymopapain (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Arg (4)	0,669935828	0,082450907
Glu (4)	0,57325757	0,038908812
His (4)	0,625279767	0,006726961
Lys (4)	0,595313259	0,009009654
Phe (4)	0,607302756	0,080934185
Ser (4)	0,737096605	0,025302226
Trp (4)	0,623929167	0,010277412
Val (4)	0,661757486	0,01771507
Ala (4)	0,579542966	0,045455096
Gln (4)	0,695667125	0,033391515
Gly (4)	0,640712756	0,008425195
Ile (4)	0,644800745	0,016665683

Appendix IX (first part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with pepsin (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Pro (1)	0,875384635	0,140316307
Thr (1)	0,749043794	0,099672962
Tyr (1)	0,758231477	0,016624087
Arg (1)	0,843318801	0,128500488
Glu (1)	0,723391778	0,013449623
His (1)	0,780989772	0,014878933
Lys (1)	0,282036156	0,017790919
Phe (1)	0,351831438	0,106808879
Ser (1)	0,818363011	0,092472501
Trp (1)	0,625432364	0,056348153
Val (1)	0,790373709	0,010197512
Ala (1)	0,449861383	0,175153351
Gln (1)	0,793969871	0,080810739
Gly (1)	0,676950355	0,087833595
Ile (1)	0,771190157	0,150367213
Pro (2)	0,784434181	0,009604871
Thr (2)	0,541201743	0,027261376
Tyr (2)	0,601176666	0,004811907
Arg (2)	0,79972906	0,092495805
Glu (2)	0,651361004	0,073782778
His (2)	0,587776824	0,067760268
Lys (2)	0,656171318	0,015841065
Phe (2)	0,804694222	0,103355683
Ser (2)	0,670587234	0,080257524
Trp (2)	0,610710133	0,067187315
Val (2)	0,710730557	0,018020322
Ala (2)	0,360275043	0,096147796
Gln (2)	0,697230959	0,090367739
Gly (2)	0,58609429	0,009829616
Ile (2)	0,689649769	0,114962705
Pro (3)	0,738729981	0,024829391
Thr (3)	0,585766235	0,059377152
Tyr (3)	0,635662114	0,061347245
Arg (3)	0,794161598	0,029426282
Glu (3)	0,62941306	0,075439116
His (3)	0,647382092	0,076930329
Lys (3)	0,618884981	0,007367057
Phe (3)	0,829874278	0,0206433
Ser (3)	0,677162882	0,087930347
Trp (3)	0,652179284	0,095379397
Val (3)	0,63315908	0,088405947
Ala (3)	0,750372628	0,004381807
Gln (3)	0,669839894	0,024253742
Gly (3)	0,621587967	0,078496183
Ile (3)	0,630348007	0,084364432
Pro (4)	0,682657144	0,068907805

Appendix IX (second part): Median values and standard errors of the ratio of fluorescein/rhodamine (five replicates) calculated after enzymatic assay with pepsin (50,625-member PNA-encoded positional scanning library). The numbers (1), (2), (3) and (4) represent the amino acids encoded in the first, second, third and fourth sub-libraries respectively.

Encoded amino acid	Median values of the ratio of fluorescein/rhodamine	Standard errors of the ratio of fluorescein/rhodamine
Thr (4)	0,584332091	0,049916419
Tyr (4)	0,699179817	0,012124259
Arg (4)	0,674457766	0,004316153
Glu (4)	0,635039888	0,082174768
His (4)	0,625258425	0,08720222
Lys (4)	0,611701579	0,085816257
Phe (4)	0,68988552	0,082587051
Ser (4)	0,684434445	0,081638005
Trp (4)	0,659044193	0,00390701
Val (4)	0,650042481	0,009429924
Ala (4)	0,632967498	0,053227016
Gln (4)	0,667135397	0,006259801
Gly (4)	0,667000262	0,004809481
Ile (4)	0,646735378	0,009360053

Appendix X: Publication

ACS chemical biology

ARTICLE

A 10,000 Member PNA-Encoded Peptide Library for Profiling Tyrosine Kinases

Delphine Pouchain, Juan J. Díaz-Mochón, Laurent Bialy, and Mark Bradley*

EaStCHEM, School of Chemistry, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3JJ, U.K.

ABSTRACT A 10,000 member peptide nucleic acid (PNA) encoded peptide library was prepared, treated with the Abelson tyrosine kinase (Abl), and decoded using a DNA microarray and a fluorescently labeled secondary antiphosphotyrosine antibody. A dual-color approach ensured internal referencing for each and every member of the library and the generation of robust data sets. Analysis identified 155 peptides (out of 10,000) that were strongly phosphorylated by Abl in full agreement with known Abl specificities. BLAST analysis identified known cellular Abl substrates such as c-Jun amino-terminal kinase as well as new potential target proteins such as the G-protein coupled receptor kinase 6 and diacylglycerol kinase gamma. To illustrate the generalization of this approach, two other tyrosine kinases, human epidermal growth factor 2 (Her2) and vascular endothelial growth factor receptor 2/kinase insert domain protein receptor (VEGFR2/KDR), were profiled allowing characterization of specific peptide sequences known to interact with these kinases; under these conditions Her2 was demonstrated to have a marked preference for D-proline perhaps offering a unique means of targeting and inhibiting this kinase.

*Corresponding author,
mark.bradley@ed.ac.uk.

Received for review September 22, 2007.
Published online December 21, 2007
10.1021/cb700199k CCC: $37.00

More than 500 kinases have been identified within the human genome, with some 90 believed to be protein tyrosine kinases, although to date only a small proportion of them have been fully characterized (*1, 2*). The uniqueness of any kinase arises from its ability to phosphorylate hydroxy residues in a highly selective and specific manner, a process by which kinases can exert exquisite and selective control over many cellular processes. Their dysfunction can be catastrophic and lead to a variety of disease states such as cancer, cardiovascular disease, inflammation, and diabetes (*3*). For instance, the breakpoint cluster region (Bcr)-Abl fusion protein has been associated with chronic myelogenous leukemia (*4, 5*). Many small-molecule kinase inhibitors such as Fasudil for Rho-dependent kinase (*6*), Sirolimus for mammalian target of rapamycin (*7*), Imatinib for Bcr-Abl (*8*), and Gefitinib for epidermal growth factor receptor (*9*) have thus emerged, but there is still a huge demand for the development of approaches that would help in the design of new kinase inhibitors (*10*).

A necessity in the area of kinase analysis is the ability to determine and understand the exact substrate specificity of any kinase, and over the past decade several kinases profiling methods have been developed. For example in 1995 a solution peptide library (prepared by split and mix methods) was used to determine the substrate specificity of nine tyrosine kinases by trapping phosphorylated peptide mixtures onto a ferric chelating column and subsequent Edman sequencing, giving a consensus peptide sequence for each kinase (*11*). A positional scanning peptide library tagged with biotin at the C-terminus, containing nine positions surrounding a central serine/threonine residue, was used to profile eight serine/threonine protein kinases, with the peptide mixtures treated in parallel with γ-[^{32}P]-ATP and a specific kinase before being immobilized onto avidin-

coated membranes allowing quantification by phosphorimaging (*12*). Another example was the use of mRNA-protein and peptide fusion libraries that were designed to have a tyrosine residue surrounded by five random amino acids to interrogate the v-abl tyrosine kinase. These libraries were designed to have one tyrosine residue surrounded by five random amino acids. Following modification of the libraries by v-abl, phosphorylated residues were isolated *via* rounds of immunoprecipitation with the phosphotyrosine-specific α4G10 antibody. Finally, amplification by PCR of the mRNA allowed DNA microarray identification (*13*). Direct bead-based assays have also allowed the identification of the substrate specificity of two protein tyrosine kinases, $p60^{c\text{-}src}$ and zeta-chain-associated protein kinase 70 (ZAP-70) (*14*). In this case, phosphorylated peptides (on beads) were identified using an antiphosphotyrosine antibody with subsequent ladder-based analysis (*15*) by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Another bead-based platform has recently been used to study kinase phosphorylation, *via* parallel peptide synthesis and subsequent encoding, with detection through chemical phosphate activation and dye coupling. Surprisingly, dyes were then detected using antibodies, presumably due to the low level of labeling (*16*). Several microarray-based methods have also been developed. Thus, self-assembled monolayers of alkanethiolates displaying peptides on gold surfaces combined with MALDI-TOF MS have been used to study kinases, although only seven peptides were studied (*17*). Microarrays of 710 peptides, generated by SPOT synthesis (*18*), with subsequent oxime attachment to a glass slide and phosphorylation in the presence of radioactive ATP allowed "phosphorimaging" and analysis of kinase specificity (*19*).

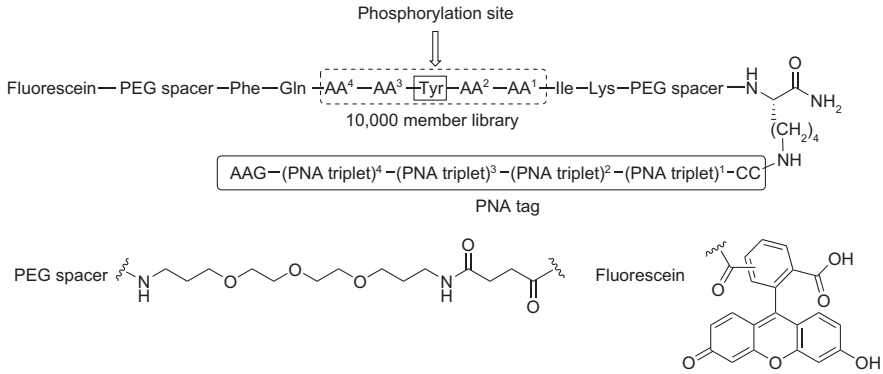
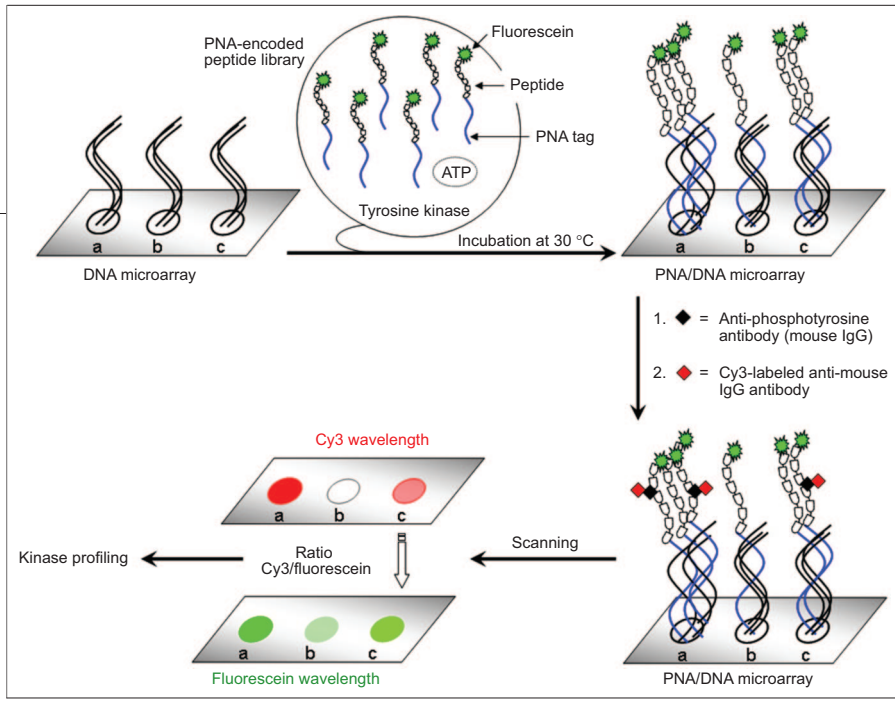


Figure 1. General structure of the 10,000 member PNA-encoded library, where AA^1, AA^2, and AA^4 are [Ile, Val, Phe, Pro, Arg, Glu, Lys, D-Pro, Ser, D-Val] and AA^3 is [Ile, Val, Ala, Pro, Arg, Glu, Lys, D-Ala, Ser, Pro].

RESULTS AND DISCUSSION

Library Design and Synthesis. Herein is described an approach that allows the screening and analysis of 10,000 kinase substrates in a single experiment, with analysis of ALL 10,000 members of the library on a one-by-one basis *via* the combined application of DNA microarrays, fluorescently labeled secondary antibodies and PNA-split and mix encoding (*20*–*23*) (the basic concept of nucleic acid encoding goes back to the early days of combinatorial chemistry (*24*)). This was achieved by the split and mix synthesis of a 10,000 member PNA-encoded library containing the peptide sequence -Phe-Gln-AA^4-AA^3-Tyr-AA^2-AA^1-Ile-Lys- (Figure 1), with the expected phosphorylation site surrounded by four variable positions (AA^1, AA^2, AA^3, and AA^4), each containing 10 different amino acids. The library contained both natural and unnatural amino acids with differing properties (hydrophobic, hydrophilic, neutral, basic, and acidic); Tyr was not included in the randomized positions to avoid multiple phosphorylation sites. Each amino acid was encoded by a specific PNA triplet to give an encoding PNA tag composed of 12 PNA monomers (Supplementary Table 1). The orthogonality and robustness of the PNA and peptide split and mix chemistry have been reported previously (*25*–*27*). In this case, four bases were used, but isothermal sequences can be generated by the use of three bases (*20, 28*). The PNA tag was also supplemented with two common PNA sequences at the C-terminus (CC) and three at the N-terminus (AAG) to help control the selectivity of hybridization and reduce mismatches (*29*). Two PEG spacers were also added between the peptide arm and the PNA oligomer to distance the peptide away from the PNA strand, and finally, every peptide was labeled with 5(6)-carboxyfluorescein in order to allow internal control of each and every library member (*20, 21*).



Scheme 1. Schematic representation of the hybridization of a PNA-encoded library incubated with the protein tyrosine kinase Abl and ATP onto a DNA microarray followed by identification of the "hits" using an antiphosphotyrosine antibody and a secondary Cy3-labeled antibody[a]
[a]Using ratios of Cy3/Fluorescein, the kinase Abl can be accurately profiled.

Abl Kinase Assay. Following solid-phase synthesis, the library was incubated with the kinase Abl (the Abelson nonreceptor tyrosine kinase that is involved in the regulation of cell proliferation, transcription, and apoptosis (*30*)) and ATP allowing phosphorylation of any peptides recognized by Abl, with subsequent pull down (hybridization) of the 10,000 member PNA-encoded library onto the DNA microarray. The arrays used contained 22,575 features with 10,000 oligonucleotide sequences complementary to the PNA sequences (each in duplicate) and 2575 control DNA sequences that were designed to be noncomplementary to any of the PNA tags in the library. In this approach, all 10,000 peptides are in essence "delivered" to a specific address on the array by virtue of their specific and unique tag (zip or postcode) *via* antiparallel PNA/DNA duplex formation (Scheme 1) (*20, 21*).

General Detection Method for Phosphorylated Peptides on Microarrays. At this stage the DNA array was converted into a peptide array containing some phosphorylated peptides that were subsequently detected using a primary/secondary antibody approach. Therefore, an antiphosphotyrosine antibody (mouse IgG) was added on the array followed by addition of a Cy3-labeled secondary antibody (antimouse IgG) (*31*). Any phosphorylated peptides ("hits"), localized onto its defined DNA, were readily identified by scanning the DNA microarray using a Cy3 filter (control experiments were carried out to prove the selectivity of the detection method using peptide microarrays, see Supporting Information). Since every peptide was also labeled with fluorescein, the resulting ratio of Cy3/Fluorescein provided an internal control for each and every member of the library, which is crucial due to natural variations in melting temperatures, concentrations, and differences in hybridization efficiency between the 10,000 different library members. The use of just Cy3 fluorescence for example would mean it would be impossible to distinguish between high-efficiency phosphorylation of a PNA–peptide conjugate that hybridizes poorly and low efficiency kinase modification with a higher concentration. Comparing the Cy3/Fluorescein ratio of every feature allows relative quantification of the fluorescence independently of any other factors and Abl can be accurately profiled, with the highest ratio of Cy3/Fluorescein corresponding to the peptide best recognized by Abl (Scheme 1).

Data Analysis. An Excel macro allowed every DNA sequence to be correlated to its PNA tag and hence its peptide, which in combination with BlueFuse (BlueGnome) allowed extensive mining and manipulation. Analysis of the images obtained after scanning the chip using a fluorescein filter showed a total absence of fluorescence on the 2575 control sequences, confirming highly selective hybridization (Figure 2). Since fluorescein was used as a label for each library member, analysis of the array using the fluorescein channel confirmed that all library members had been synthesized and hybridized onto the DNA microarray. In order to quantify the level of phosphorylation, a second channel (Cy3) was used (Scheme 1). Although similarities with gene expression profiling exist, the analysis used here differed in many respects due to the fact that the majority of the features in the Cy3 channel had little if any signal (most peptides were nonphosphorylated). In order to gain valuable data, only features (specific peptides) with a Cy3 intensity twice that of the background were considered (giving some 703 initial hits). These data were then verified by identification of all duplicates from within these 703 single signals (duplicates of each DNA oligomer existed scattered across the array) removing those with a standard deviation higher than 0.25 (comparing normalized ratios of Cy3/Fluorescein intensities for each fea-

ture with values ranging from 0 to 1). In doing so, 155 peptides ("310 duplicative hits") gave data for Abl, with very high confidence (see Supporting Information for details). The same analysis approach was used for the two other kinases (see Supplementary Figure 2 for Abl profiles and Supplementary Figure 3 for the image of the chip obtained using a Cy3 filter).

Specificity of Abl. Broad specificity for the amino acids in positions AA^4 and AA^1 was found whereas a high selectivity was clearly observed for the amino acids around the phosphorylation site (Tyr) in the AA^3 and AA^2 positions. Glu seems to play a crucial role in Abl phosphorylation since it was the most frequent amino acid found in each randomized position, confirming that this protein tyrosine kinase accepts acidic residues around the phosphorylation site (Asp was not chosen as a building block in the library synthesis due to its similarity to Glu) (*11, 14*). Glu (85 "hits" out of 155, 55%) was clearly the most accepted amino acid in the AA^3 position. Thus, an acidic, negatively charged amino acid on the amino side of tyrosine seems to be important for efficient phosphorylation by Abl.Smaller proportions of Val (18%) and Ser (11%) were also observed. In the AA^2 position, Abl displayed a preference for polar amino acids such as Ser (35%) and acidic residues such as Glu (32%). In the AA^4 and AA^1 positions, Abl was observed to be less specific and accepted a broad range of amino acids such as Ser (24%), Glu (19%), D-Val (17%) in the AA^4 position and Ser (22%), Val (18%), Glu (13%) in the AA^1 position (Figure 3). When these results were compared with the sequences of known Abl kinase substrates, marked similarities were displayed (Table 1).

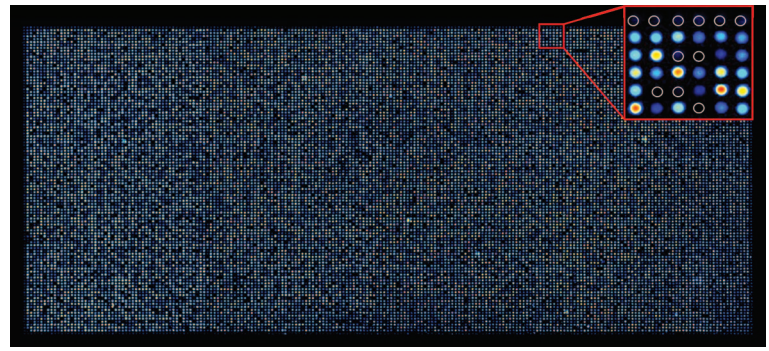


Figure 2. Image of a 10,000 member PNA-encoded kinase library hybridized onto a 22,575 member custom DNA microarray, following incubation with Abl, obtained using a fluorescein filter. Every fluorescent spot represents a unique peptide, whose sequence is known via its PNA tag and location on the array. Empty white circles (insert) correspond to DNAs that were designed to be noncomplementary to any of the PNA tags in the library and were printed as controls of the hybridization process. Note the variation in fluorescence intensity which highlights differences in hybridization and synthesis efficiency that have to be controlled using relative ratios of fluorescence rather than absolute values.

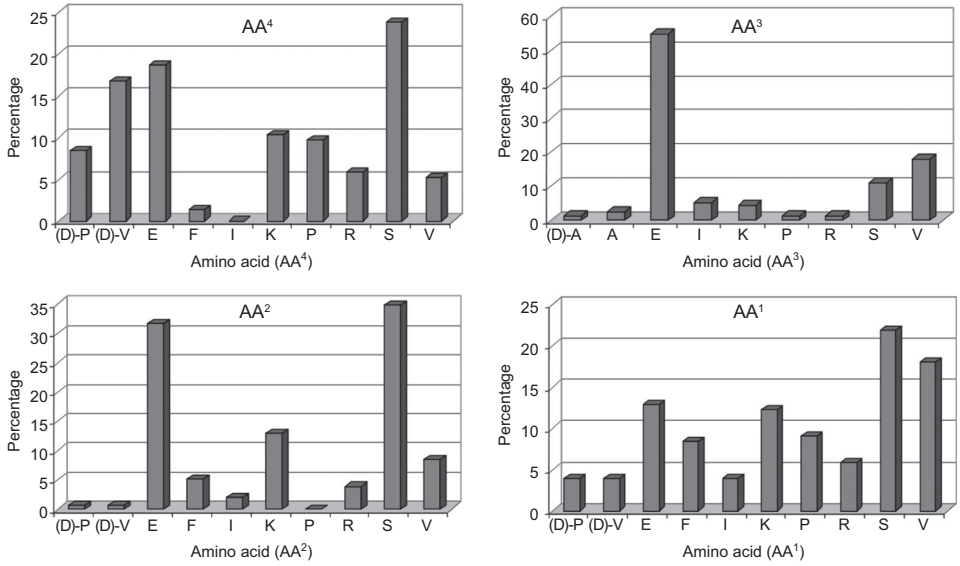


Figure 3. Bar graphs for Abl showing the proportion of amino acids at the four randomized positions (AA^4, AA^3, AA^2, and AA^1) in the 155 "hits".

TABLE 1. Comparison between the sequences of known Abl substrates and the most frequent amino acids obtained from the library screening (*32*)[a]

Substrate	Swiss-Prot	Position	Sequence
SHP1	PTN6_HUMAN (MOUSE)	Y536	**SEY**GN
SHP1	PTN6_HUMAN (MOUSE)	Y564	**DVYE**N (**EVYE**N)
Mdm2	MDM2_HUMAN (MOUSE)	Y394 (Y393)	**EDYS**Q (**DDYS**Q)
p62dok	DOK1_MOUSE	Y314	**SVYSD**
p62dok	DOK1_MOUSE	Y397	**E**G**YE**L
Myogenic factor 3	MYOD1_MOUSE	Y30	**DFYDD**
Myogenic factor 3	MYOD1_MOUSE	Y212	M**DYS**G

[a]The known tyrosine phosphorylation site is indicated and bold letters represent amino acids that were common to our study (one-letter amino acid codes are used).

Specificity of the Top Peptide Sequences for Abl. The data from the 155 "hits" could also be analyzed to determine the best substrates accepted by Abl, and the sequences of the top 10 peptides are listed in Table 2. The motif SEY appeared 5 times (out of 10), revealing this to be an important factor in Abl specificity. Moreover, among the 155 "hits", 21 peptides contained the SEY motif and 14 were found in the top 40 peptide sequences (Supplementary Table 2). The Abl optimal substrate predicted by Songyang (*11*) contains the motif I/VYXXP, where X represents any amino acid. Although Pro is replaced by Ile in our library (Figure 1), 8 peptides contained Ile (5%, Figure 3) in the AA^3 position and they were among those giving the highest ratios of Cy3/Fluorescein (Supplementary Figure 2). Thus in the top 10 peptide sequences, Ile appeared three times in the AA^3 position (Table 2).

Interacting Partners for Abl. A BLAST search using the Swiss-Prot database was performed with the top 25 peptide sequences identified (*32*, *33*). The results were observed to be in agreement with known Abl signaling pathways and proteins implicated in leukemia (*30*, *34*–*36*). Thus, the second best peptide sequence (SEYES) was the well-known Abl substrate, c-Jun amino-terminal kinase (Table 3) (*34*). Although the role of Abl with G protein coupled receptor kinases and diacylglycerol kinase is not fully understood, our data reveal that Abl may interact with these proteins (*37*, *38*).

Screening Additional Kinases. To further demonstrate the validity of the method, two additional kinases (Her2 and VEGFR2/KDR) were screened using the same 10,000 member PNA-encoded peptide library, giving two new profiles (Supplementary Figures 4 and 5). Her2 has been found to be amplified in more than 25% of human primary breast cancers (*39*) and VEGFR2/KDR plays crucial roles in angiogenesis and vasculogenesis (*40*). Surprisingly, Her2 showed a remarkable preference for the unnatural amino acid D-Pro at the C-terminus site of the phosphorylation site. Thus, looking at the 299 "hits", in the AA^1 position, D-Pro was the most accepted amino acid (34%) along with Ser (28%). This was also observed in the AA^2 position, where D-Pro (22%), Ser (20%), and D-Val (17%) were the most common amino acids. Moreover, in the top 15 peptides (Supplementary Table 3), 10 peptides in the AA^1 position and 7 in the AA^2 position contained D-Pro, with the sequence D-Pro/D-Pro (AA^2/AA^1) represented five times. These results were particularly interesting since even though L-Pro was present in positions AA^1 and AA^2, Her2 had a clear preference for D-Pro. The marked preference for this unnatural amino acid was unexpected and would not be observed in other library screens. However, it is already known that protein kinase C can phosphorylate both D- and L-amino acids (*41*). Her2 was less specific for the amino acids in positions AA^4 and AA^3, but a small preference for Lys (18%), Glu (14%), and Ser (14%) in the AA^4 position and Ser (23%), Lys

TABLE 2. Sequences of the top 10 peptides, recognized by Abl, with the highest ratios of Cy3/Fluorescein[a]

Peptide sequence	Ratio Cy3/Fluorescein	Peptide sequence	Ratio Cy3/Fluorescein
1) SEYEV	4.50	6) SIYEP	3.44
2) SEYES	4.45	7) SIYSP	3.32
3) SEYSF	4.16	8) PEYSE	3.29
4) SEYVF	4.03	9) SEYEE	3.27
5) VIYES	3.62	10) VEYES	3.26

[a]The one letter amino acid code is used.

TABLE 3. Predicted proteins for Abl identified by running a BLAST search using peptide sequences from the 25 top peptides and the Swiss-Prot database (only proteins with the highest scores were selected)

Protein name	Swiss-Prot	Identified sequence
G-protein coupled receptor kinase	GRK6_MOUSE	SEYEV
c-Jun N terminal kinase	JIP2_MOUSE	SEYES
Diacylglycerol kinase gamma	DGKG_MOUSE	SEYSS
G-protein coupled receptor kinase	GRK5_MOUSE	EVYSS
C−C chemokine receptor	CCR6_MOUSE	RVYSE
Epidermal growth factor receptor substrate	EP15_MOUSE	PVYEK

(17%), and D-Ala (16%) in the AA^3 position could be noticed (Supplementary Figure 4). Unlike the two other kinases, VEGFR2/KDR was less specific and accepted a broad range of amino acids, giving a much greater number of "hits" (829 "hits") (Supplementary Figure 5). A slight preference for polar amino acids (both acidic and basic) immediately adjacent to the phosphorylation site (positions AA^3 and AA^2) could be observed: Glu (AA^3/AA^2, 20%/16%), Lys (AA^3/AA^2, 17%/16%), and Ser (AA^3/AA^2, 17%/14%) were the most commonly observed amino acids. In the AA^4 and AA^1 positions, Lys (15%) and D-Pro (18%), respectively, were the most common amino acids.

Interacting Partners for Her2 and VEGFR2/KDR. A BLAST search using the Swiss-Prot database was performed with the top 25 peptide sequences of Her2 and VEGFR2/KDR (Table 4, Supplementary Tables 3 and 4). Although only six peptide sequences (within the top 25) contained natural amino acids for Her2, proteins that are known to interact with Her2 or present in breast cancer could be identified (*42–44*). For VEGFR2/KDR, the results were also in agreement with known signaling proteins (*45–47*).

CONCLUSION

The high-throughput screening of 10,000 tyrosine kinase Abl substrates in a single experiment has been demonstrated *via* the detection of tyrosine-phosphorylated peptides ("hits") using a nonradioactive, selective and sensitive method based on a primary antiphosphotyrosine antibody and a fluorescently labeled secondary antibody. The use of PNA-encoded libraries simplified the deconvolution process while using minimal amounts of kinase (60 pmol), library (5 nmol), and antibodies (0.4 μg per antibody). ALL 10,000 members of the library were analyzed, allowing a unique insight into the substrate requirements of Abl; however it is a general method as there is no need to have any previous knowledge about an opti-

TABLE 4. Predicted protein targets for Her2 and VEGFR2/KDR identified by running a BLAST search using peptide sequences from the 25 top peptides identified from the microarray and the Swiss-Prot database (only proteins with the highest scores were selected)[a]

Tyrosine kinase	Protein name (sequence)	Swiss-Prot	Identified sequence
Her2	Breast tumor-amplified kinase (YQETYKRI)	STK6_HUMAN	F**QE**I**YKRI**
Her2	G protein coupled receptor 158 (EIYKRKK)	GP158_ HUMAN	**EIYKR**I**K**
Her2	Ephrin receptor EPHA7 (FQTRYPS)	EPHA7_ HUMAN	**FQ**S**RYPS**
VEGFR2/KDR	Leptin receptor isoform 1 (FQIRY)	LEPR_HUMAN	**FQIRY**
VEGFR2/KDR	Protein kinase C, theta (IVYRD)	KPCT_ HUMAN	**IVYR**E
VEGFR2/KDR	Endothelial differentiation G protein coupled receptor 6 (IIYSF)	EDG6_ HUMAN	**IIYSF**

[a]Bold letters for the identified peptide sequence represent amino acids that are common to the predicted protein.

mal kinase substrate. The dual color approach (ratio of Cy3/Fluorescein) was essential to allow accurate profiling and gave a set of peptides that were phosphorylated in agreement with known Abl interacting partners, while new target proteins were quickly and simply identified by analysis of the top peptide sequences and BLAST searching. Moreover, it is known that Abl accepts a wide range of substrates and one of the advantages of our method is that optimal peptides for Abl, rather than consensus sequences, can be identified. Finally, unlike other methods (*14*) that give large number of "hits" without relative quantification, the approach described here offers the advantage of not only detecting phosphorylated peptides but also quantifying the extent of phosphorylation. The approach was used to profile and identify specific proteins for two other tyrosine kinases, Her2 and VEGFR2/KDR, showing its generality, while adaptations (*16*) will allow serine and threonine kinases to be analyzed allowing rapid and efficient substrate deorphaning.

METHODS

Materials. Fmoc amino acids were purchased from GL Biochem except Fmoc-Tyr(HPO_3Bzl)-OH and Fmoc-Ser(HPO_3Bzl)-OH which were from LC Sciences. CodeLink activated slides were obtained from Amersham Biosciences. ATP and monoclonal antiphosphotyrosine clone PY-20 were from Sigma, Cy3-goat antimouse IgG (H+L) conjugate (ZyMax grade) was from Invitrogen. Abl, mouse recombinant (*Escherichia coli*) was from New England Biolabs, Her2 (human, recombinant, N-terminal GST tag) and VEGFR2/KDR (human, recombinant, N-terminal GST tag) were purchased from BPS Bioscience.

Library Synthesis. The PNA-encoded peptide library (see Figure 1 for general structure of the library and Supplementary Table 1 for PNA code) was synthesized according to published procedures (*25, 26*). It was carried out on PEGA resin with a Rink linker where Fmoc-PEG spacer-Lys(Dde)-OH [PEG spacer = NH-$(CH_2)_3$-O-$(CH_2)_2$-O-$(CH_2)_2$-O-$(CH_2)_3$-NH-CO-$(CH_2)_2$-CO] was attached as the core unit presenting two different orthogonal protecting groups (Fmoc = 20% piperidine in DMF and Dde = hydroxylamine, pH 5.0). This orthogonality allowed elongation of either the peptide or PNA arm at will through selective Fmoc and Dde chemistries.

Thus, Fmoc-Rink-amide linker was attached to PEGA resin following stepwise couplings with Fmoc-Lys(Dde)-OH (*48*), Fmoc-PEG spacer-OH (*49*), Fmoc-Lys(Boc)-OH, and Fmoc-Ile-OH. Then, the Dde group was removed (NH_2OH.HCl/imidazole in NMP/DCM, 1 h) and two cytosine PNA monomers were coupled to the resin. The synthesis was then carried out as follows: (1) Split into 10 different pools. (2) Dde deprotection. (3) Dde/Mmt PNA monomer coupling: Dde-PNA(Mmt)-OH (5.5 equiv) and PyBOP (5 equiv) were dissolved in DMF (0.1 M) followed by addition of NEM (11 equiv). The resulting solution was mixed for 10 s before being added to the resin preswollen in DMF, and the mixture was shaken for 3 h. Repeat (2) and (3) twice. (4) Fmoc deprotection (20% piperidine in DMF, two cycles of 10 min). (5) Fmoc amino acids couplings: Fmoc-AA-OH (5.5 equiv), PyBOP (5 equiv), and HOBt (5.5 equiv) were dissolved in DMF (0.08 M) followed by addition of DIPEA (16 equiv). The resulting solution was mixed for 10 s before being added to the resin, and the mixture was shaken for 3 h. (6) The 10 resin pools were mixed. (7) The resin was split again into 10 pools; repeat steps (2) to (6). After encoding two positions and before splitting, Fmoc-Tyr(tBu)-OH was coupled to the resin. Then the synthesis continued (steps (1)−(7)). At the end of the encoding process, Fmoc-Gln(Trt)-OH, Fmoc-Phe-OH, and Fmoc-PEG spacer-OH were respectively coupled followed by coupling of guanine, adenine, and adenine PNA-protected monomers at the PNA arm as previously described. Finally, the library was labeled using 5(6)-carboxyfluorescein (Fluorescein) (5.5 equiv), PyBOP (5 equiv), and NEM (5 equiv) in DMF (0.1 M). Before cleavage, the library was treated with 20% piperidine in DMF (5 × 10 min) (*50*). Final cleavage from the resin was carried out using TFA/TIS/DCM (90/5/5) for 1 h. The crude compound was precipitated from cold diethyl ether, sonicated for 5 min, and centrifuged for 5 min at 4000 rpm. The supernatant was then discarded, and the process was repeated four times to give rise to a yellow solid.

Solution-Phase Kinase Assay. The 10,000 member PNA-encoded library was incubated, in solution, at a final concentration of 50 μM and final volume of 100 μL, in a buffer containing 60 units of the selected kinase (Abl, Her2, or VEGFR2/KDR), 5 mM ATP, kinase reaction buffer (for Abl 1x, 50 mM Tris-HCl (pH 7.5 at 25 °C), 10 mM $MgCl_2$, 1 mM EGTA, 2 mM DTT, 0.01% Brij 35; or for Her2 and VEGFR2/KDR, 50 mM HEPES, pH 7.4, 3 mM $MgCl_2$, 3 mM $MnCl_2$, 1 mM DTT, 3 μM Na-orthovanadate) and 0.1% BSA for 24 h at 30 °C. Before hybridization, 1 volume of this buffer was diluted with 1 volume of GenHyb buffer (Genetix). Solutions were denatured by heating at 65 °C for 3 min; 100 μL of the solution was then poured onto a 22,500 customized DNA chip from Oxford Gene Technology (Oxford), which was covered with a glass cover slide and kept in a hybridization chamber (Genomic Solutions) at 60 °C. Temperature was then slowly lowered to 37 °C over a period of 21 h. Finally, the chips were washed twice (20 mL) for 10 min at 30 °C in the following buffer (100 mM Nacl, 10 mM citric acid, 0.7% (w/v) *N*-lauroylsarcosine sodium salt, 0.1 mM EGTA, pH 7.5), rinsed with distilled water (20 mL), and spin-dried by centrifugation (1000 rpm for 10 min) (*20*).

Antibody Assay on the 22,500 Customized DNA Chip. To avoid any unspecific binding, the slides were blocked with a prewarmed buffer containing TPBS (0.05% Tween-20/PBS, pH 7.4)/1% BSA (20 mL) for 1 h at 30 °C. After being washed in PBS (20 mL) for 5 min at 30 °C, slides, that were covered with a glass cover-slide and kept in a hybridization chamber (Genomic Solutions), were probed with the primary antibody, monoclonal antiphosphotyrosine clone PY-20 (mouse IgG2b isotype), diluted 1/500 (0.4 μg) in an antibody buffer containing TPBS/1% BSA (final volume of 200 μL) for 1 h at 30 °C. After the slides were washed with the antibody buffer (2 × 20 mL) for 10 min at 30 °C, a labeled secondary antibody, Cy3-goat antimouse IgG (H+L), diluted 1/500 (0.2 μg) in the antibody buffer (final volume of 200 μL) was added and slides were incubated in the dark for 1 h at 30 °C. Slides were then washed in the antibody buffer (2 × 20 mL for 10 min at 30 °C) and rinsed with Tris buffer (20 mL) at 25 °C before being spin-dried by centrifugation (1000 rpm for 10 min) (*31*).

Acknowledgment: This work was supported by the EPSRC/BBSRC and the Combinatorial Center of Excellence (CCE).

Supporting Information Available: This material is free of charge *via* the Internet.

REFERENCES

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome, *Science 298*, 1912–1934.
2. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling, *Nature 411*, 355–365.
3. Cohen, P. (2002) Protein kinases-the major drug targets of the twenty-first century? *Nat. Rev. Drug Discovery 1*, 309–315.
4. Rowley, J. D. (1990) The philadelphia chromosome translocation a paradigm for understanding leukemia, *Cancer 65*, 2178–2184.
5. Till, J. H., Chan, P. M., and Miller, W. T. (1999) Engineering the substrate specificity of the Abl tyrosine kinase, *J. Biol. Chem. 274*, 4995–5003.
6. Asano, T., Ikegaki, I., Satoh, S.-I., Seto, M., and Sasaki, Y. (1998) A protein kinase inhibitor, Fasudil (AT-877): a novel approach to signal transduction therapy, *Cardiovasc. Drug Rev. 16*, 76–87.
7. Sehgal, S. N. (1998) Rapamune (RAPA, rapamycin, sirolimus): mechanism of action immunosuppressive effect results from blockade of signal transduction and inhibition of cell cycle progression, *Clin. Biochem. 31*, 335–340.
8. Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., and Lydon, N. B. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells, *Nat. Med. 2*, 561–566.
9. Barker, A. J., Gibson, K. H., Grundy, W., Godfrey, A. A., Barlow, J. J., Healy, M. P., Woodburn, J. R., Ashton, S. E., Curry, B. J., Scarlett, L., Henthorn, L., and Richards, L. (2001) Studies leading to the identification of ZD1839 (Iressa): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer, *Bioorg. Med. Chem. Lett. 11*, 1911–1914.
10. Noble, M. E. M., Endicott, J. A., and Johnson, L. N. (2004) Protein kinase inhibitors: insights into drug design from structure, *Science 303*, 1800–1805.
11. Songyang, Z., Carraway III, K. L., Eck, M. J., Harrison, S. C., Feldman, R. A., Mohammadi, M., Schlessinger, J., Hubbard, S. R., Smith, D. P., Eng, C., Lorenzo, M. J., Ponder, B. A. J., Mayer, B. J., and Cantley, L. C. (1995) Catalytic specificity of protein-tyrosine kinases is critical for selective signalling, *Nature 373*, 536–539.
12. Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., Cantley, L. C., and Turk, B. E. (2004) A rapid method for determining protein kinase phosphorylation specificity, *Nat. Methods 1*, 27–29.
13. Cujec, T. P., Medeiros, P. F., Hammond, P., Rise, C., and Kreider, B. L. (2002) Selection of v-Abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display, *Chem. Biol. 9*, 253–264.
14. Kim, Y.-G., Shin, D.-S., Kim, E.-M., Park, H.-Y., Lee, C.-S., Kim, J.-H., Lee, B.-S., Lee, Y.-S., and Kim, B.-G. (2007) High-throughput identification of substrate specificity for protein kinase by using an improved one-bead-one-compound library approach, *Angew. Chem., Int. Ed. 46*, 5408–5411.
15. Pastor, J. J., Lingard, I., Bhalay, G., and Bradley, M. (2003) Ion-extraction ladder sequencing from bead-based libraries, *J. Comb. Chem. 5*, 85–90.
16. Shults, M. D., Kozlov, I. A., Nelson, N., Kermani, B. G., Melnyk, P. C., Shevchenko, V., Srinivasan, A., Musmacker, J., Hachmann, J. P., Barker, D. L., Lebl, M., and Zhao, C. (2007) A multiplexed protein kinase assay, *ChemBioChem 8*, 933–942.
17. Min, D.-H., Su, J., and Mrksich, M. (2004) Profiling kinase activities by using a peptide chip and mass spectrometry, *Angew. Chem., Int. Ed. 43*, 5973–5977.
18. Frank, R. (1992) Spot-synthesis an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support, *Tetrahedron 48*, 9217–9232.
19. Schutkowski, M., Reimer, U., Panse, S., Dong, L., Lizcano, J. M., Alessi, D. R., and Schneider-Mergener, J. (2004) High-content peptide microarrays for deciphering kinase specificity and biology, *Angew. Chem., Int. Ed. 43*, 2671–2674.
20. Diaz-Mochon, J. J., Bialy, L., Keinicke, L., and Bradley, M. (2005) Combinatorial libraries - from solution to 2D microarrays, *Chem. Commun. 11*, 1384–1386.
21. Diaz-Mochon, J. J., Bialy, L., and Bradley, M. (2006) Dual colour, microarray-based, analysis of 10 000 protease substrates, *Chem. Commun. 38*, 3984–3986.
22. Winssinger, N., Harris, J. L., Backes, B. J., and Schultz, P. G. (2001) From split-pool libraries to spatially addressable microarrays and its application to functional proteomic profiling, *Angew. Chem., Int. Ed. 40*, 3152–3155.
23. Winssinger, N., Damoiseaux, R., Tully, D. C., Geierstanger, B. H., Burdick, K., and Harris, J. L. (2004) PNA-encoded protease substrate microarrays, *Chem. Biol. 11*, 1351–1360.
24. Brenner, S., and Lerner, R. A. (1992) Encoded combinatorial chemistry, *Proc. Natl. Acad. Sci. U.S.A. 89*, 5381–5383.
25. Bialy, L., Diaz-Mochon, J. J., Specker, E., Keinicke, L., and Bradley, M. (2005) Dde-protected PNA monomers, orthogonal to Fmoc, for the synthesis of PNA-peptide conjugates, *Tetrahedron 61*, 8295–8305.
26. Diaz-Mochon, J. J., Bialy, L., and Bradley, M. (2004) Full orthogonality between Dde and Fmoc: the direct synthesis of PNA-peptide conjugates, *Org. Lett. 6*, 1127–1129.
27. Diaz-Mochon, J. J., Bialy, L., Watson, J., Sanchez-Martin, R. M., and Bradley, M. (2005) Synthesis and cellular uptake of cell delivering PNA-peptide conjugates, *Chem. Commun. 26*, 3316–3318.
28. Brenner, S., Williams, S. R., Vermaas, E. H., Storck, T., Moon, K., McCollum, C., Mao, J.-I, Luo, S., Kirchner, J. J., Eletr, S., DuBridge, R. B., Burcham, T., and Albrecht, G. (2000) In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs, *Proc. Natl. Acad. Sci. U.S.A. 97*, 1665–1670.
29. Weiler, J., Gausepohl, H., Hauser, N., Jensen, O. N., and Hoheisel, J. D. (1997) Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays, *Nucleic Acids Res. 25*, 2792–2799.
30. Pendergast, A. M. (2002) The Abl family kinases: mechanisms of regulation and signaling, *Adv. Cancer Res. 85*, 51–100.
31. Horiuchi, K. Y., Wang, Y., Diamond, S. L., and Ma, H. (2006) Microarrays for the functional analysis of the chemical-kinase interactome, *J. Biomol. Screen. 11*, 48–56.
32. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res. 31*, 365–370.
33. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389–3402.
34. Yoshida, K., Yamaguchi, T., Natsume, T., Kufe, D., and Miki, Y. (2005) JNK phosphorylation of 14–3-3 proteins regulates nuclear targeting of c-Abl in the apoptotic response to DNA damage, *Nat. Cell Biol. 7*, 278–285.
35. Dürig, J., Schmücker, U., and Dührsen, U. (2001) Differential expression of chemokine receptors in B cell malignancies, *Leukemia 15*, 752–756.
36. Tanos, B., and Pendergast, A. M. (2006) Abl tyrosine kinase regulates endocytosis of the epidermal growth factor receptor, *J. Biol. Chem. 281*, 32714–32723.

37. Ushio-Fukai, M., Zuo, L., Ikeda, S., Tojo, T., Patrushev, N. A., and Alexander, R. W. (2005) cAbl tyrosine kinase mediates reactive oxygen species- and caveolin-dependent AT 1 receptor signaling in vascular smooth muscle: role in vascular hypertrophy, *Circ. Res. 97*, 829–836.
38. Yamada, K., Sakane, F., Imai, S.-I., Tsushima, S., Murakami, T., and Kanoh, H. (2003) Regulatory role of diacylglycerol kinase γ in macrophage differentiation of leukemia cells, *Biochem. Biophys. Res. Commun. 305*, 101–107.
39. Hayes, D. F., Thor, A. D., Dressler, L. G., Weaver, D., Edgerton, S., Cowan, D., Broadwater, G., Goldstein, L. J., Martino, S., Ingle, J. N., Henderson, I. C., Norton, L., Winer, E. P., Hudis, C. A., Ellis, M. J., and Berry, D. A. (2007) Her2 and response to paclitaxel in node-positive breast cancer, *N. Engl. J. Med. 357*, 1496–1506.
40. Olsson, A.-K, Dimberg, A., Kreuger, J., and Claesson-Welsh, L. (2006) VEGF receptor signalling-in control of vascular function, *Nat. Rev. Mol. Cell Biol. 7*, 359–371.
41. Eller, M., Jarv, J., Toomik, R., Ragnarsson, U., Ekman, P., and Engstrom, L. (1993) Substrate specificity of protein kinase C studied with peptides containing D-amino acid residues, *J. Biochem. 114*, 177–180.
42. Sen, S., Zhou, H., and White, R. A. (1997) A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines, *Oncogene 14*, 2195–2200.
43. Negro, A., Brar, B K, Gu, Y., Peterson, K. L., Vale, W., and, and Lee, K.-F (2006) ErbB2 is required for G protein-coupled receptor signaling in the Heart, *Proc. Natl. Acad. Sci. U.S.A. 103*, 15889–15893.
44. Fox, B. P., and Kandpal, R. P. (2004) Invasiveness of breast carcinoma cells and transcript profile: Eph receptors and ephrin ligands as molecular markers of potential diagnostic and prognostic application, *Biochem. Biophys. Res. Commun. 318*, 882–892.
45. Gonzalez, R. R., Cherfils, S., Escobar, M., Yoo, J. H., Carino, C., Styer, A. K., Sullivan, B. T., Sakamoto, H., Olawaiye, A., Serikawa, T., Lynch, M. P., and Rueda, B. R. (2006) Leptin signaling promotes the growth of mammary tumors and increases the expression of vascular endothelial growth factor (VEGF) and its receptor type two (VEGF-R2), *J. Biol. Chem. 281*, 26320–26328.
46. Singh, A. J., Meyer, R. D., Band, H., and Rahimi, N. (2005) The carboxyl terminus of VEGFR-2 is required for PKC-mediated down-regulation, *Mol. Biol. Cell 16*, 2106–2118.
47. Endo, A., Nagashima, K.-I., Kurose, H., Mochizuki, S., Matsuda, M., and Mochizuki, N. (2002) Sphingosine 1-phosphate induces membrane ruffling and increases motility of human umbilical vein endothelial cells via vascular endothelial growth factor receptor and CrkII, *J. Biol. Chem. 277*, 23747–23754.
48. Chhabra, S. R., Hothi, B., Evans, D. J., White, P. D., Bycroft, B. W., and Chan, W. C. (1998) An appraisal of new variants of Dde amine protecting group for solid phase peptide synthesis, *Tetrahedron Lett. 39*, 1603–1606.
49. Zhao, Z. G., Im, J. S., Lam, K. S., and Lake, D. F. (1999) Site-specific modification of a single-chain antibody using a novel glyoxylyl-based labeling reagent, *Bioconjugate Chem. 10*, 424–430.
50. Fischer, R., Mader, O., Jung, G., and Brock, R. (2003) Extending the applicability of carboxyfluorescein in solid-phase synthesis, *Bioconjugate Chem. 14*, 653–660.